

TAB DD

[54] METHOD OF CONSTRUCTING A
REPLICABLE CLONING VEHICLE HAVING
QUASI-SYNTHETIC GENES[75] Inventors: David V. Goeddel; Herbert L.
Heyneker, both of Burlingame, Calif.[73] Assignee: Genentech, Inc., South San
Francisco, Calif.

[21] Appl. No.: 55,126

[22] Filed: Jul. 5, 1979

[51] Int. Cl.³ C12N 15/00[52] U.S. Cl. 435/172; 435/68;
435/70; 435/317; 536/27[58] Field of Search 435/317, 172, 68, 70,
435/71

[56] References Cited

FOREIGN PATENT DOCUMENTS

6694 6/1979 European Pat. Off. .
20147 12/1980 European Pat. Off. .

OTHER PUBLICATIONS

- Technology Review, pp. 12 and 13, Dec. 1976.
 Martial et al., Science, vol. 205, Aug. 10, 1979.
 Shine et al., Nature, vol. 285, Jun. 12, 1980, pp. 456-461.
 The Economist, pp. 87 and 88, Jul. 14, 1979.
 Time, Jul. 30, 1970, p. 70.
 Newmark, Nature, vol. 280, pp. 637 and 638, Aug. 23, 1979.
 Villa-Komaroff et al., Proc. Natl. Acad. Sci., vol. 75, pp. 3727-3731, Aug. 1978.
 Seeburg et al., Nature, vol. 276, pp. 795-798, Dec. 1978.
 Itakura et al., Science, vol. 198, pp. 1056-1063, Dec. 1977.
 Crea et al., Proc. Natl. Acad. Sci., vol. 75, pp. 5765-5769, Dec. 1978.

Klenow et al., Proc. Natl. Acad. Sci., vol. 65, pp. 168-175, Jan. 1970.

Sutcliffe, Cold Spring Harbor Symposium 43, pp. 70-90 (1978).

Curtis et al., Molecular Cloning of Recombinant DNA, by Scott et al., pp. 99-111 (1977).

Ullrich et al., Science, vol. 196, pp. 1313-1319, Jun. 1977.

Bolivar et al., Gene 2, pp. 95-113 (1977).

Goeddel et al., Proc. Natl. Acad. Sci., vol. 76, pp. 106-110, Jan. 1979.

Chang et al., Nature, vol. 275, pp. 617-624, Oct. 1978.

Maxam et al., Proc. Nat. Acad. Sci., vol. 74, pp. 560-564 (Feb. 1977).

Kornberg, DNA Synthesis, pp. 87 and 88, pub. by W. H. Freeman & Co., 1974.

Razin et al., Proc. Natl. Acad., vol. 75, pp. 4268-4270, Sep. 1978.

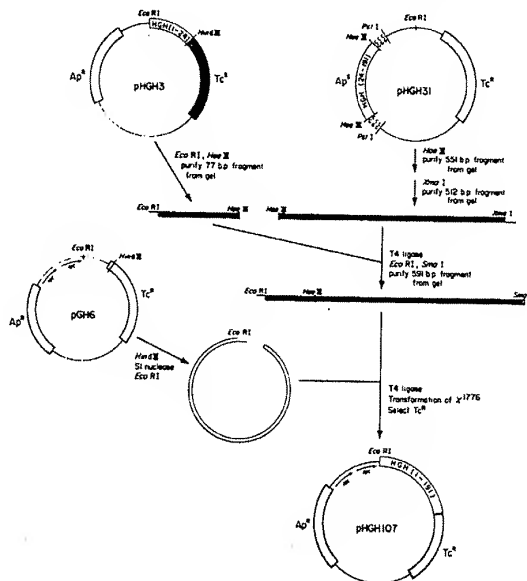
Wickens et al., The Journal of Biological Chemistry, vol. 253, No. 7, pp. 2483-2495 (1978).

Primary Examiner—Alvin E. Tanenholzt
Attorney, Agent, or Firm—Thomas D. Kiley

[57] ABSTRACT

Described are methods and means for the construction and microbial expression of quasi-synthetic genes arising from the combination of organic synthesis and enzymatic reverse transcription from messenger RNA sequences incomplete from the standpoint of the desired protein product. Preferred products of expression lack bio-inactivating leader sequences common in eukaryotic expression products but problematic with regard to microbial cleavage to yield bioactive material. Illustrative is a preferred embodiment in which a gene coding for human growth hormone (useful in, e.g., treatment of hypopituitary dwarfism) is constructed and expressed.

12 Claims, 5 Drawing Figures



HGH (1-24)

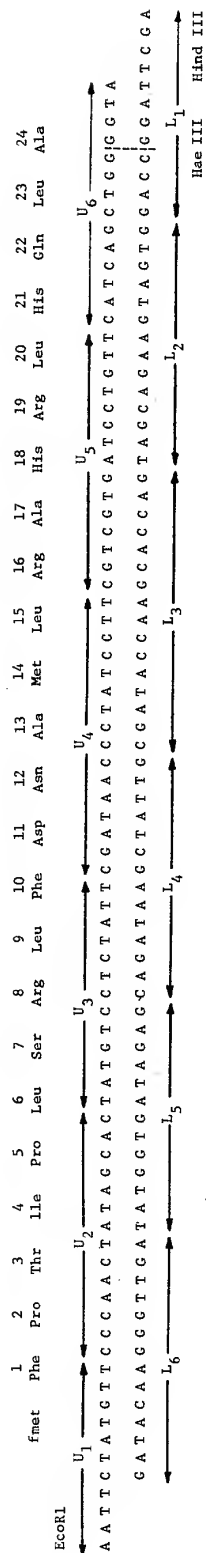


Fig. 1.

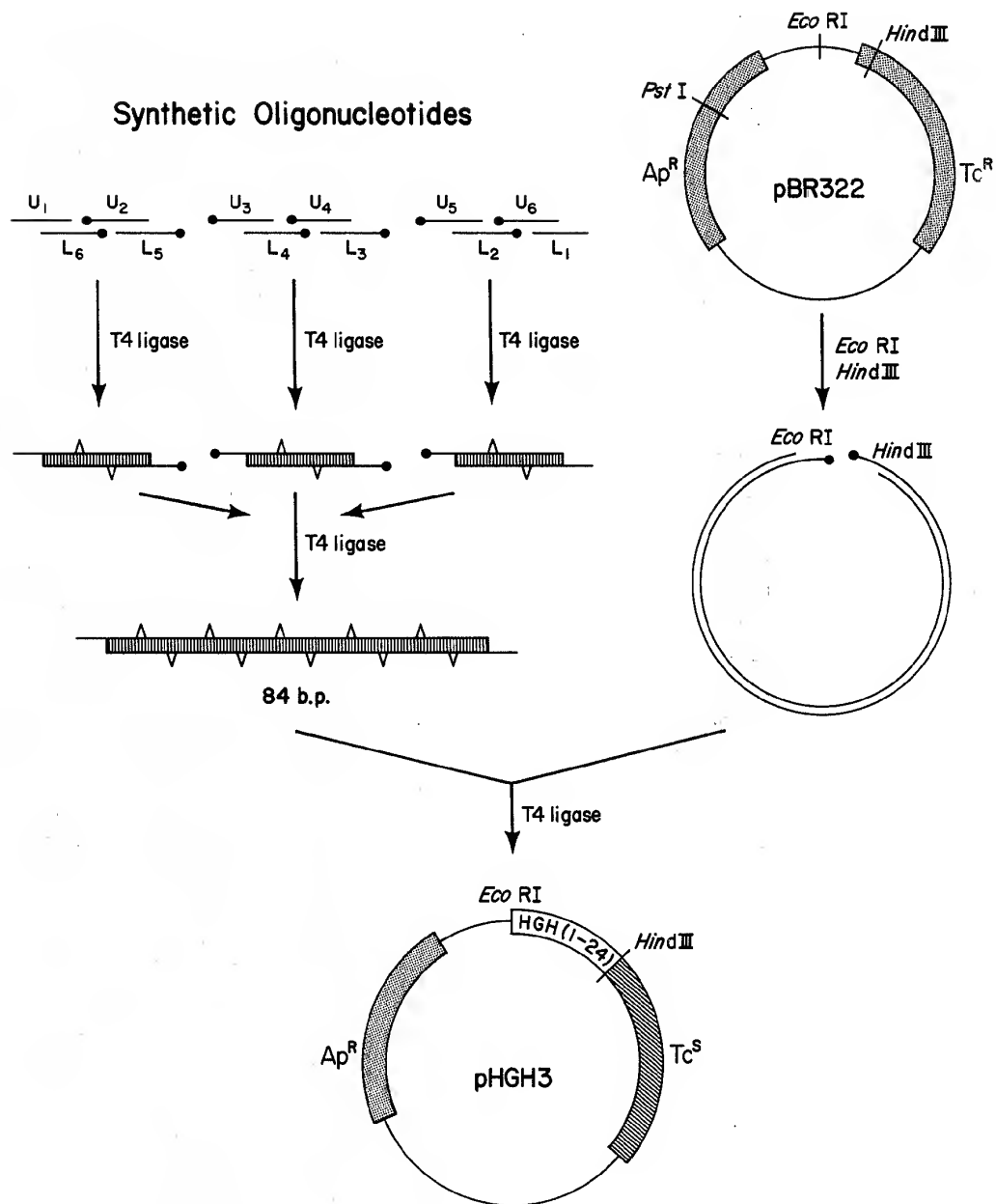


FIG. 2.

24
 Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn Pro Gln Thr Ser
 5'...G GCC UUU GAC ACC GAG GAG UUU GAA GAA GCC UAU AUC CCA AAG GAA CAG AAG UAU UCA UUC CUG CAG AAC CCC CAG ACC UCC
 40
Pst I
 Leu Cys Phe Ser Glu Ser Ile Pro Thr Tyr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Arg Ile Ser Leu Leu
 60
 CUC UGU UUC UCA GAG UCU AUU CCG ACA CCG UCC AAC AGG GAG GAA ACA CAA CAG AAA UCC AAC CUA GAG CUG CUC CGC AUC UCC CUG CUG
 80
 100
 Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp Ser Asn Val Tyr
 CUC AUC CAG UCG UGG CUG GAG CCC GUG CAG UUC CUC AGG AGU GUC UUC GCC AAC AGC CUA GUG UAC GGC GCC UCU GAC AGC AAC GUC UAU
 120
 Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe Lys Gln
 GAC CUC CUA AAG GAC CUA GAG GAA GGC AUC CAA ACG CUG AUG GGG AGG CUG GAA GAU GGC AGC CCC CGG ACU GGG CAG AUC UUC AAG CAG
 140
Bgl II
 Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp
 160
 ACC UAC AGC AAG UUC GAC ACA AAC UCA CAC AAC GAA GAC GCA CUA CUC AAG AAC UAC GGC CUG CUC UAC UGC UUC AGG AAG GAC AUG GAC
 180
 Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly Ser Cys Gly Phe Stop
 AAG GUC GAG ACA UUC CUG CGC AUC GUG CAG UGC CGC UCU GUG GAG GGC AGC UGU GGC UUC UAG CUGCCCCGGUGGCAUCCUGUGAGCCCCUCCCGAGU
 191
Pvu II
Sma I
Xma I
 GCCUCUGCCGCG...3'
Hae III

Fig. 3.

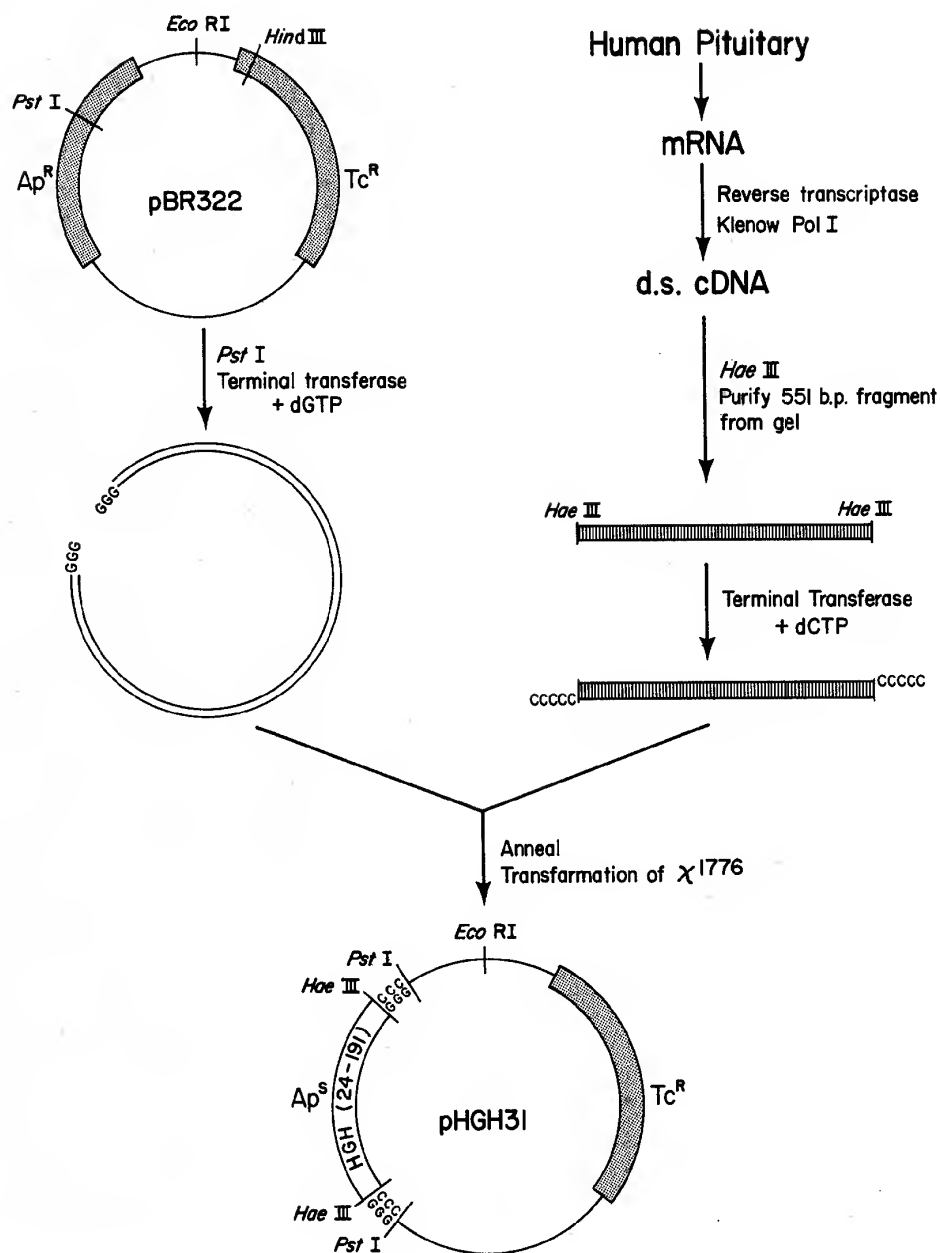


FIG. 4.

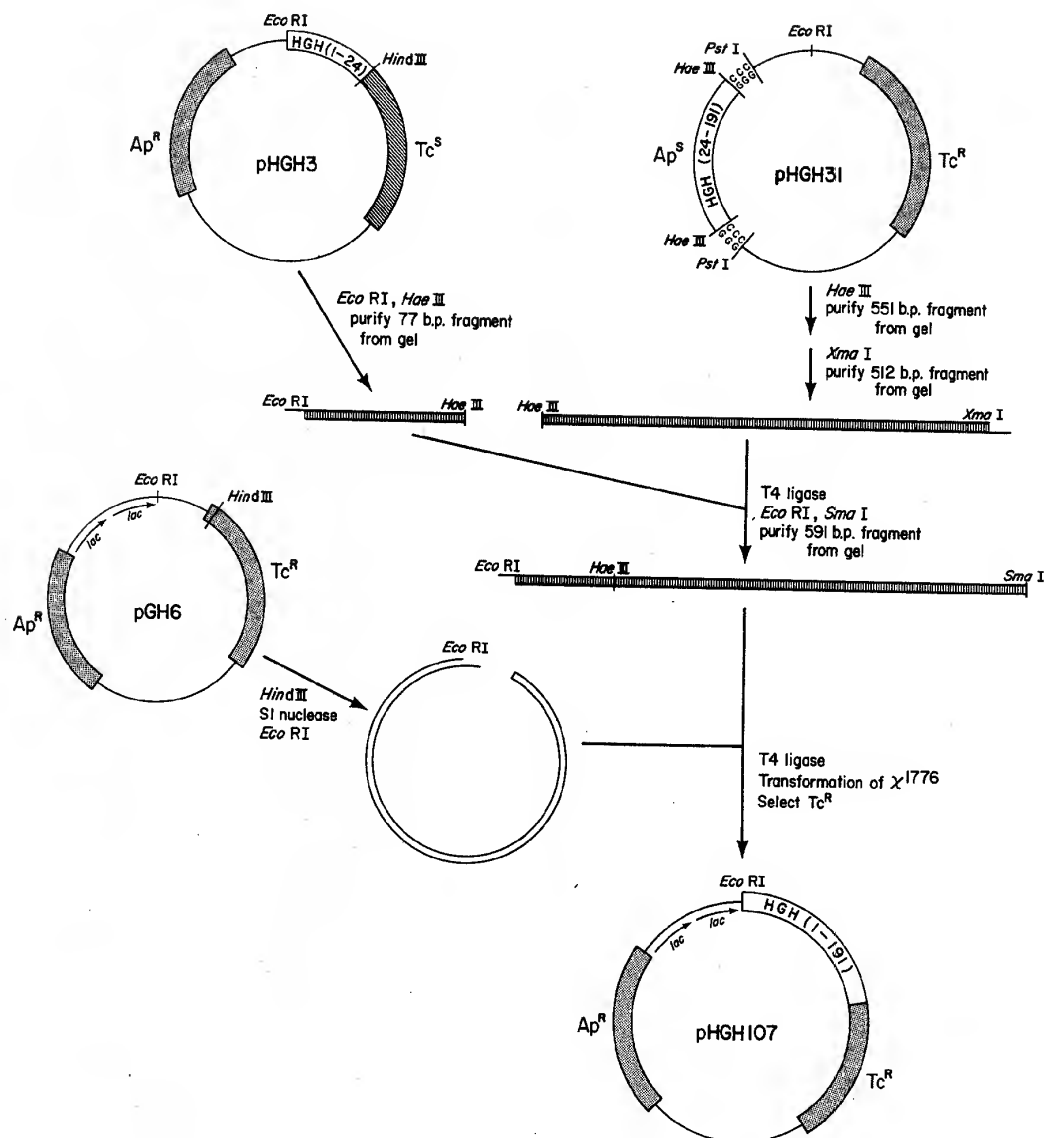


FIG. 5.

METHOD OF CONSTRUCTING A REPLICABLE CLONING VEHICLE HAVING QUASI-SYNTHETIC GENES

BACKGROUND OF THE INVENTION

Genetic Expression

The DNA (deoxyribonucleic acid) of which genes are made comprises both protein-encoding or "structural" genes and control regions that mediate the expression of their information through provision of sites for RNA polymerase binding, information for ribosomal binding sites, etc. Encoded protein is "expressed" from its corresponding DNA by a multistep process within an organism by which:

1. The enzyme RNA polymerase is activated in the control region (hereafter the "promoter") and travels along the structural gene, transcribing its encoded information into messenger ribonucleic acid (mRNA) until transcription of translatable mRNA is ended at one or more "stop" codons.

2. The mRNA message is translated at the ribosomes into a protein for whose amino acid sequence the gene encodes, beginning at a translation "start" signal, most commonly ATG (which is transcribed "AUG" and translated "f-methionine").

In accordance with the genetic code, DNA specifies each amino acid by a triplet or "codon" of three adjacent nucleotides individually chosen from adenosine, thymidine, cytidine and guanine or, as used herein, A,T,C, or G. These appear in the coding strand or coding sequence of double-stranded ("duplex") DNA, whose remaining or "complementary" strand is formed of nucleotides ("bases") which hydrogen bond to their complements in the coding strand. A complements T, and C complements G. These and other subjects relating to the background of the invention are discussed at length in Benjamin Lewin, *Gene Expression* 1, 2 (1974) and 3 (1977), John Wiley and Sons, N.Y. This and the other publications alluded to herein are incorporated by reference.

DNS CLEAVAGE AND LIGATION

A variety of techniques are available for DNA recombination, according to which adjoining ends of separate DNA fragments are tailored in one way or another to facilitate ligation. The latter term refers to the formation of phosphodiester bonds between adjoining nucleotides, most often through the agency of the enzyme T4 DNA ligase. Thus, blunt ends may be directly ligated. Alternatively, fragments containing complementary single strands at their adjoining ends are advantaged by hydrogen bonding which positions the respective ends for subsequent ligation. Such single strands, referred to as cohesive termini, may be formed by the addition of nucleotides to blunt ends using terminal transferase, and sometimes simply by chewing back one strand of a blunt end with an enzyme such as λ -exonuclease. Again, and most commonly, resort may be had to restriction endonucleases (hereafter, "restriction enzymes"), which cleave phosphodiester bonds in and around unique sequences of nucleotides of about 4-6 base pairs in length ("restriction sites"). Many restriction enzymes and their recognition sites are known. See, e.g., R. J. Roberts, *CRC Critical Reviews in Biochemistry*, 123 (November 1976). Many make staggered cuts that generate short complementary single-stranded sequences at the ends of the duplex fragments. As comple-

mentary sequences, the protruding or "cohesive" ends can recombine by base pairing. When two different molecules are cleaved with this enzyme, crosswise pairing of the complementary single strands generates a new DNA molecule, which can be given covalent integrity by using ligase to seal the single strand breaks that remain at the point of annealing. Restriction enzymes which leave coterminal or "blunt" ends on duplex DNA that has been cleaved permit recombination via, e.g., T4 ligase with other blunt-ended sequences.

CLONING VEHICLES AND RECOMBINANT DNA

For present purposes, a "cloning vehicle" is an extra-chromosomal length of duplex DNA comprising an intact replicon such that the vehicle can be replicated when placed within a unicellular organism ("microbe") by transformation. An organism so transformed is called a "transformant". Presently, the cloning vehicles commonly in use are derived from viruses and bacteria and most commonly are loops of bacteria DNA called "plasmids".

Advances in biochemistry in recent years have led to the construction of "recombinant" cloning vehicles in which, for example, plasmids are made to contain exogenous DNA. In particular instances the recombinant may include "heterologous" DNA, by which is meant DNA that codes for polypeptides ordinarily not produced by the organism susceptible to transformation by the recombinant vehicle. Thus, plasmids are cleaved with restriction enzymes to provide linear DNA having ligatable termini. These are bound to an exogenous gene having ligatable termini to provide a biologically functional moiety with an intact replicon and a phenotypical property useful in selecting transformants. The recombinant moiety is inserted into a microorganism by transformation and the transformant is isolated and cloned, with the object of obtaining large populations that include copies of the exogenous gene and, in particular cases, with the further object of expressing the protein for which the gene codes. The associated technology and its potential applications are reviewed in extenso in the Miles International Symposium Series 10: *Recombinant Molecules: Impact on Science and Society*, Beers and Bosseff, eds., Raven Press, N.Y. (1977).

RECOMBINANT DNA EXPRESSION

Aside from the use of cloning vehicles to increase the supply of genes by replication, there have been attempts, some successful, to actually express proteins for which the genes code. In the first such instance a gene for the brain hormone somatostatin under the influence of the lac promoter was expressed in *E. Coli* bacteria. K. Itakura et al., *Science* 198, 1056 (1977). More recently, the A and B chains of human insulin were expressed in the same fashion and combined to form the hormone. D. V. Goeddel et al., *Proc. Nat'l. Acad. Sci., U.S.A.* 76, 106 (1979). In each case the genes were constructed in their entirety by synthesis. In each case, proteolytic enzymes within the cell would apparently degrade the desired product, necessitating its production in conjugated form, i.e., in tandem with another protein which protected it by compartmentalization and which could be extracellularly cleaved away to yield the product intended. This work is described in the following published British patent specifications of the assignee of the present application: GB No. 2,007,675 A;

GB No. 2,007,670 A; GB No. 2,007,676 A; and GB No. 2,008,123 A.

While the synthetic gene approach has proven useful in the several cases thus far discussed, real difficulties arise in the case of far larger protein products, e.g., growth hormone, interferon, etc., whose genes are correspondingly more complex and less susceptible to facile synthesis. At the same time, it would be desirable to express such products unaccompanied by conjugate protein, the necessity of whose expression requires diversion of resources within the organism better committed to construction of the intended product.

Other workers have attempted to express genes derived not by organic synthesis but rather by reverse transcription from the corresponding messenger RNA purified from tissue. Two problems have attended this approach. To begin with, reverse transcriptase may stop transcription from mRNA short of completing cDNA for the entire amino acid sequence desired. Thus, for example, Villa-Komaroff et al obtained cDNA for rat proinsulin which lacked codons for the first three amino acids of the insulin precursor. *Proc. Nat'l. Acad. Sci., U.S.A.* 75 3727 (1978). Again, reverse transcription of mRNA for polypeptides that are expressed in precursor form has yielded cDNA for the precursor form rather than the bioactive protein that results when, in a eukaryotic cell, leader sequences are enzymatically removed. Thus far, no bacterial cell has been shown to share that capability, so that mRNA transcripts have yielded expression products containing the leader sequences of the precursor form rather than the bioactive protein itself. Villa-Komaroff, supra (rat proinsulin); P. H. Seeburg et al., *Nature* 276, 795 (1978) (rat pregrowth hormone).

Finally, past attempts by others to bacterially express hormones (or their precursors) from mRNA transcripts have on occasion led only to the production of conjugated proteins not apparently amenable to extra-cellular cleavage, e.g., Villa-Komaroff, supra, (penicillinase-proinsulin); Seeburg, supra (beta-lactamase-pregrowth hormone).

HUMAN GROWTH HORMONE

Human growth hormone ("HGH") is secreted in the human pituitary. It consists of 191 amino acids and, with its molecular weight of about 21,500, is more than three times as large as insulin. Until the present invention, human growth hormone could be obtained only by laborious extraction from a limited source—the pituitary glands of human cadavers. The consequent scarcity of the substance has limited its applications to the treatment of hypopituitary dwarfism, and even here reliable estimates suggest that human-derived HGH is available in sufficient quantity to serve not more than about 50% of afflicted subjects.

In summary, a need has existed for new methods of producing HGH and other polypeptide products in quantity, and that need has been particularly acute in the case of polypeptides too large to admit of organic synthesis or convenient synthesis of genes from which the peptide could be expressed. Expression of mammalian hormones from mRNA transcripts has offered the promise of side-stepping difficulties that attend the synthetic approach, but until the present has permitted only microbial production of bio-inactive conjugates from which the desired hormone could not practicably be cleaved.

SUMMARY OF THE INVENTION

The present invention provides methods and means for expressing quasi-synthetic genes wherein reverse transcription provides a substantial portion, preferably a majority, of the coding sequence without laborious resort to entirely synthetic construction, while synthesis of the remainder of the coding sequence affords a completed gene capable of expressing the desired polypeptide unaccompanied by bio-inactivating leader sequences or other extraneous protein. Alternatively, the synthetic remainder may yield a proteolysis-resistant conjugate so engineered as to permit extra-cellular cleavage of extraneous protein, yielding the bioactive form. The invention accordingly makes available methods and means for microbial production of numerous materials hitherto produced only in limited quantity by costly extraction from tissue, and still others previously incapable of industrial manufacture. In its most preferred embodiment the invention represents the first occasion in which a medically significant polypeptide hormone (human growth hormone) has been bacterially expressed while avoiding both intracellular proteolysis and the necessity of compartmentalizing the bioactive form in extraneous protein pending extracellular cleavage. Microbial sources for human growth hormone made available by the invention offer, for the first time, ample supplies of the hormone for treatment of hypopituitary dwarfism, together with other applications heretofore beyond the capacity of tissue-derived hormone sources, including diffuse gastric bleeding, pseudarthrosis, burn therapy, wound healing, dystrophy and bone knitting.

The manner in which these and other objects and advantages of the invention may be obtained will appear more fully from the detailed description which follows, and from the accompanying drawings relating to a preferred embodiment of the invention, in which:

FIG. 1 depicts the synthetic scheme for construction of a gene fragment coding for the first 24 amino acids of human growth hormone, together with the start signal ATG and linkers used in cloning. The arrows in the coding or upper strand ("U") and in the complementary or lower strands ("L") indicate the oligonucleotides joined to form the depicted fragment;

FIG. 2 depicts joinder of the "U" and "L" oligonucleotides to form the gene fragment of FIG. 1, and its insertion in a plasmid cloning vehicle;

FIG. 3 illustrates the DNA sequence (coding strand only) of the Hae III restriction enzyme fragment of a pituitary mRNA transcript, with the numbered amino acids of human growth hormone for which they code. Key restriction sites are indicated, as is DNA (following "stop") for untranslated mRNA;

FIG. 4 illustrates the construction of a cloning vehicle for a gene fragment coding for the amino acids of human growth hormone not synthetically derived, and the construction of that gene fragment as complementary DNA by reverse transcription from mRNA isolated from a human pituitary source; and

FIG. 5 illustrates the construction of a plasmid capable, in bacteria, of expressing human growth hormone, beginning with the plasmids of FIGS. 2 and 4.

DETAILED DESCRIPTION OF THE INVENTION

The general approach of the invention involves the combination in a single cloning vehicle of plural gene

fragments which in combination code for expression of the desired product. Of these, at least one is a cDNA fragment derived by reverse transcription from mRNA isolated from tissue, as by the method of A. Ullrich et al., *Science* 196, 1313 (1977). The cDNA provides a substantial portion, and preferably at least a majority, of the codons for the desired product, while remaining portions of the gene are supplied synthetically. The synthetic and mRNA transcript fragments are cloned separately to provide ample quantities for use in the later combination step.

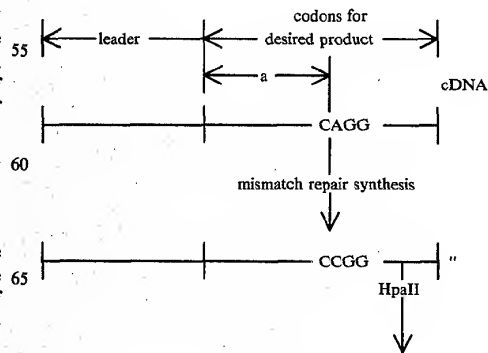
A variety of considerations influence distribution of codons for the end product as between synthetic and cDNA, most particularly the DNA sequence of complementary DNA determined as by the method of Maxam and Gilbert, *Proc. Nat'l Acad. Sci. U.S.A.* 74, 560 (1977). Complementary DNA obtained by reverse transcription will invariably contain codons for at least a carboxy terminal portion of the desired product, as well as other codons for untranslated mRNA downstream from the translation stop signal(s) adjacent the carboxy terminus. The presence of DNA for untranslated RNA is largely irrelevant, although unduly lengthy sequences of that kind may be removed, as by restriction enzyme cleavage, to conserve cellular resources employed in replicating and expressing the DNA for the intended product. In particular cases, the cDNA will contain codons for the entire amino acid sequence desired, as well as extraneous codons upstream from the amino terminus of the intended product. For example, many if not all polypeptide hormones are expressed in precursor form with leader or signal sequences of protein involved, e.g., in transport to the cellular membrane. In expression from eukaryotic cells, these sequences are enzymatically removed, such that the hormone enters the periplasmic space in its free, bioactive form. However, microbial cells cannot be relied upon to perform that function, and it is accordingly desirable to remove sequences coding for such signals or leader sequences from the mRNA transcript. In the course of that removal process the translation start signal is also lost, and almost invariably some codons for the intended product will be removed as well. The synthetic component of the quasi-synthetic gene product of the invention returns these latter codons, as well as supplying anew a translation start signal where the vehicle into which the hybrid gene will ultimately be deployed itself lacks a properly positioned start.

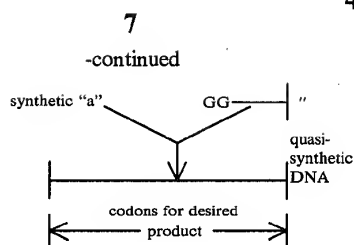
Elimination of the leader sequence from pregrowth hormone cDNA is advantaged by the availability of a restriction site within the growth hormone-encoding portion of the gene. The invention may nevertheless be practiced without regard to the availability of such a site, or in any event without regard to the availability of a restriction site sufficiently near the amino terminus of the desired polypeptide as to obviate the need for extensive synthesis of the gene component not derived from mRNA. Thus, in any cDNA coding for the desired polypeptide and a leader or other bioactivating sequence the boundary between the latter's codons and those of the mature polypeptide will appear from the amino acid sequence of the mature polypeptide. One may simply digest into the gene coding of the peptide of choice, removing the unwanted leader or other sequence. Thus, for example, given cDNA such as:



where the endpoint of digestion is indicated by arrow, reaction conditions for exonuclease digestion may be chosen to remove the upper sequences "a" and "b", whereafter S1 nuclease digestion will automatically eliminate the lower sequences "c" and "d". Alternatively and more precisely, one may employ DNA polymerase digestion in the presence of deoxynucleotide triphosphates ("d(A,T,C,G)TP"). Thus, in the foregoing example, DNA polymerase in the presence of dGTP will remove sequence "c" (then stop at "G"), S1 nuclease will then digest "a", DNA polymerase in the presence of dTTP will remove "d", (then stop at "T") and S1 nuclease will then excise "b", and so on. See generally A. Kornberg, *DNA Synthesis*, pp. 87-88, W. H. Freeman and Co., San Francisco (1974).

More preferably, one may simply construct a restriction site at a convenient point within the portion of the cDNA coding for the desired product, by an application of the mismatch repair synthesis technique of A. Razin et al., *Proc. Nat'l Acad. Sci. USA* 75, 4268 (1978). By this technique one or more bases may be substituted in an existing DNA sequence, using primers containing the mismatched substituent. At least seven palindromic 4-base pair sequences are uniquely recognized by known restriction enzymes, i.e., AGCT (Alu I), CCGG (Hpa II), CGCG (Tha I), GATC (Sau 3A), GCGC (Hha), GGCC (Hae III), and TCGA (Taq I). Where the cDNA sequence contains a sequence differing from one such site in a single base, as statistically is highly likely, repair synthesis will yield replicate cDNA containing the proper, substituent base and hence the desired restriction site. Cleavage will delete DNA for the unwanted leader, after which synthesis will replace codons required for expression of the complete polypeptide. E.g.,:





It will be appreciated, of course, that longer restriction sites may be likewise intended where desired, or that successive repairs may create 4-base pair restriction sites where only two bases common to the site appear at the desired point, etc.

Applications will appear in which it is desirable to express not only the amino acid sequence of the intended product, but also a measure of extraneous but specifically engineered protein. Four such applications may be mentioned by way of example. First, the quasi-synthetic gene may represent a hepten or other immunological determinant upon which immunogenicity is conferred by conjugation to additional protein, such that vaccines are produced. See generally, G. B. patent specification No. 2,008,123A. Again, it may be desirable for bio-safety reasons to express the intended product as a conjugate with other, bio-inactivating protein so designed as to permit extracellular cleavage to yield the active form. Third, applications will be presented in which transport signal polypeptides will precede the desired product, to permit production of the same by excretion through the cell membrane, so long as the signal peptide can then be cleaved. Finally, extraneous conjugate designed to permit specific cleavage extracellularly may be employed to compartmentalize intended products otherwise susceptible to degradation by proteases endogenous to the microbial host. At least in the latter three applications, the synthetic adaptor molecular employed to complete the coding sequence of the mRNA transcript can additionally incorporate codons for amino acid sequences specifically cleavable, as by enzymatic action. For example, trypsin will cleave specifically at arg-arg or lys-lys, etc. See GB 2 008 123A, supra.

From the foregoing, it will be seen that in its broadest aspect the invention admits of manifold applications, each having in common these attributes:

a mRNA transcript is employed which codes for a substantial portion of the intended polypeptide's amino acid sequence but which, if expressed alone, would produce a different polypeptide either smaller or larger than the intended product;

protein-encoding codons for amino acid sequences other than those contained in the intended product, if any, are removed;

organic synthesis yields fragment(s) coding for the remainder of the desired sequence; and

the mRNA transcript and synthetic fragment(s) are combined and disposed in a promoter-containing cloning vehicle for replication and expression of either the intended product absent extraneous conjugated protein, or intended product conjugated to but specifically cleavable from extraneous protein.

Of course, the expression product will in every case commence with the amino acid coded for by the translation start signal (in the case of ATG, f-methionine). One can expect this to be removed intracellularly, or in

any event to leave the bioactivity of the ultimate product essentially unaffected.

Although it provides a method of general applicability in the production of useful proteins, including antibodies, enzymes and the like, the invention is particularly suited to the expression of mammalian polypeptide hormones and other substances having medical applications, e.g., glucagon, gastrointestinal inhibitory polypeptide, pancreatic polypeptide, adrenocorticotropin, beta-endorphins, interferon, urokinase, blood clotting factors, human albumin, and so on. A preferred embodiment illustrative of the invention is next discussed, in which a quasi-synthetic gene coding for human growth hormone is constructed, cloned and microbially expressed.

CONSTRUCTION AND EXPRESSION OF A CLONING VEHICLE FOR HUMAN GROWTH HORMONE

1. Cloning the Hae III fragment of the mRNA transcript (FIGS. 3 and 4)

Polyadenylated mRNA for human growth hormone (HGH) was prepared from pituitary growth hormone-producing tissue by the procedure of A. Ullrich et al. *Science* 196, 1313 (1977). 1.5 μ g of double strand ("ds") cDNA was prepared from 5 μ g of this RNA essentially as described by Wickens et al. *J. Biol. Chem.* 253 2483 (1978), except that RNA polymerase "Klenow fragment", H. Klenow, *Proc. Nat'l. Acad. Sci. USA* 65, 168 (1970), was substituted for DNA Polymerase I in the second strand synthesis. The restriction pattern of HGH is such that Hae III restriction sites are present in the 3' non-coding region and in the sequence coding for amino acids 23 and 24 of HGH, as shown in FIG. 3. Treatment of ds HGH cDNA with Hae III gives a DNA fragment of 551 base pairs ("bp") coding for amino acids 24-191 of HGH. Thus, 90 ng of the cDNA was treated with Hae III, electrophoresed on an 8% polyacrylamide gel, and the region at 550 bp eluted. Approximately 1 ng of cDNA was obtained.

pBR322 prepared as in F. Bolivar et al., *Gene* 2 (1977) 95-113 was chosen as the cloning vehicle for the cDNA. pBR322 has been fully characterized, J. G. Sutcliffe, *Cold Spring Harbor Symposium* 43, 70 (1978), is a multicopy replicating plasmid which exhibits both ampicillin and tetracycline resistance owing to its inclusion of the corresponding genes ("Ap^R" and "Tc^R", respectively, in FIG. 4), and which contains recognition sites for the restriction enzymes Pst I, EcoRI and Hind III as shown in the Figure.

The GC tailing method of Chang, A. C. Y. et al. *Nature* 275 617 (1978) was employed to combine the products of Pst I cleavage of pBR322 and of Hae III digestion of the mRNA transcript, inserting the cDNA fragment into the Pst I site of pBR322 in such manner as to restore the Hae III restriction sites (GG \downarrow CC) on the cDNA while restoring the Pst I restriction sites (CTGCA \downarrow G) at each end of the insert.

Thus, terminal deoxynucleotidyl transferase (TdT) was used to add approximately 20 dC residues per 3' terminus as described previously, Chang, A. Y. C., supra. 60 ng of Pst I-treated pBR322 was tailed similarly with about 10 dG residues per 3' terminus. Annealing of the dC-tailed ds cDNA with the dG-tailed vector DNA was performed in 130 μ l of 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.25 mM EDTA. The mixture was heated to 70° C., allowed to cool slowly to 37° C. (12

hours), then to 20° C. (6 hours) before being used to transform *E. Coli*. ×1776. DNA sequence analysis of the plasmid pHGH31 cloned in ×1776 by the method of Maxam and Gilbert, *Proc. Nat'l. Acad. Sci. USA* 74, 560 (1977) resulted in confirmation of the codons for amino acids 24-191 of HGH, as shown in FIG. 3.

E. Coli K-12 strain ×1776 has the genotype F⁻ tonA53 dapD8 minA1 supE42 Δ40[gal-uvrB] λ⁻ minB2 rfb-2 nalA25 oms-2 thyA57* metC65 oms-1 Δ29[bioH-asd] cycB2 cycA1 hsdR2. ×1776 has been certified by the National Institutes of Health as an EK2 host vector system.

×1776 has an obligate requirement for diaminopimelic acid (DAP) and cannot synthesize the mucopolysaccharide colanic acid. It thus undergoes DAP-less death in all environments where DAP is limiting but sufficient nutrients exist to support cellular metabolism and growth. It requires thymine or thymidine and undergoes thymineless death with degradation of DNA when thymine and thymidine are absent from the environment but when sufficient nutrients are present to sustain metabolic activity. ×1776 is extremely sensitive to bile and thus is unable to survive and thus is unable to survive passage through the intestinal tract of rats. ×1776 is extremely sensitive to detergents, antibiotics, drugs and chemicals. ×1776 is unable to carry out either dark or photo repair of UV-induced damage and is thus several orders of magnitude more sensitive to sunlight than wild-type strains of *E. Coli*. ×1776 is resistant to many transducing phages and is conjugation deficient for inheritance of many different types of conjugative plasmids due to the presence of various mutations. ×1776 is resistant to nalidixic acid, cycloserine and trimethoprim. These drugs can therefore be added to media to permit monitoring of the strain and to preclude transformation of contaminants during transformation.

×1776 grows with a generation time of about 50 min. in either L broth or Penassay broth when supplemented with 100 μg DAP/ml and 4 μg thymidine/ml and reaches final densities of 8-10×10⁸ cells/ml at stationary phase. Gentle agitation by swirling and shaking back and forth for a period of 1-2 min. adequately suspends cells with maintenance of 100% viability. Additional details concerning ×1776 appear in R. Curtis et al., *Molecular Cloning of Recombinant DNA*, 99-177, Scott and Werner, eds., Academic Press (New York 1977).

2. Construction and Cloning of the Synthetic Gene Fragment (FIGS. 1 and 2)

The strategy for construction of the HGH quasisynthetic gene included construction of a synthetic fragment comprising a blunt-end restriction cleavage site adjacent the point at which the fragment would be joined to the mRNA transcript. Thus, as shown in FIG. 1, the synthetic gene for the first 24 amino acids of HGH contained a Hae III cleavage site following amino acid 23. The distal end of the synthetic fragment was provided with a "linker" that permitted annealing to a single strand terminal resulting from restriction cleavage in the plasmid in which the mRNA transcript and synthetic fragment would ultimately be joined.

As shown in FIG. 1, the 5' ends of the duplex fragment have single stranded cohesive termini for the Eco RI and Hind III restriction endonucleases to facilitate plasmid construction. The methionine codon at the left end provides a site for initiation of translation. Twelve different oligonucleotides, varying in size from un-

decamer to hexadecamer, were synthesized by the improved phosphotriester method of Crea, R. *Proc. Nat'l. Acad. Sci. USA* 75, 5765 (1978). These oligonucleotides, U₁ to U₆ and L₁ to L₆, are indicated by arrows.

10 μg amounts of U₂ through U₆ and L₂ through L₆ were phosphorylated using T₄ polynucleotide kinase and (γ³²-P)ATP by a published procedure. Goeddel, D. V. et al. *Proc. Nat'l. Acad. Sci. USA* 76, 106 (1979).

Three separate T₄ ligase catalyzed reactions were performed: 10 μg of 5'-OH fragment U₁ was combined with the phosphorylated U₂, L₅ and L₆; phosphorylated U₃, U₄, L₃ and L₄ were combined; and 10 μg of 5'-OH fragment L₁ was combined with the phosphorylated L₂, U₅ and U₆. These ligations were carried out at 4° C. for 6 hours in 300 μl of 20 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM ATP using 20 units of T₄ ligase. The three ligation mixtures were then combined, 20 units T₄ ligase added, and the reaction allowed to proceed for 12 hours at 20° C. The mixture was ethanol precipitated and electrophoresed on a 10% polyacrylamide gel. The band migrating at 84 base pairs was sliced from the gel and eluted. pBR322 (1 μg) was treated with Eco RI and Hind III, the large fragment isolated by gel electrophoresis and ligated to the synthetic DNA. This mixture was used to transform *E. Coli* K-12 strain 294 (end A, thi⁻, hsr⁻, hsm_k⁺). Strain 294 was deposited Oct. 30, 1978 in the American Type Culture Collection (ATCC No. 31446), without restriction. Sequence analysis by the Maxam and Gilbert technique, supra, on the Eco RI-Hind III insert from a plasmid pHGH3 of one transformant confirmed that depicted in FIG. 1.

3. CONSTRUCTION OF PLASMID FOR THE BACTERIAL EXPRESSION OF HGH (FIG. 5)

With the synthetic fragment in pHGH3 and the mRNA transcript in pHGH31, a replicable plasmid containing both fragments was constructed using the expression plasmid pGH6, as shown in FIG. 5. The expression plasmid, which contains tandem lac promoters, was first constructed as follows. A 285 base pair Eco RI fragment containing two 95 base pair UV5 lac promoter fragments separated by a 95 base pair heterologous DNA fragment was isolated from plasmid pKB268, K. Backman, et al., *Cell*, Vol. 13, 65-71 (1978). The 285 bp fragment was inserted into the Eco RI site of pBR322 and a clone pGH1 isolated with the promoters oriented toward and in proper reading phase with the gene for tetracycline resistance. The Eco RI site distal to the latter gene was destroyed by partial Eco RI digestion, repair of the resulting single stranded Eco RI ends with DNA polymerase I and recircularization of the plasmid by blunt-end ligation. The resulting plasmid, pGH6, contains a single Eco RI site properly positioned with respect to the promoter system into which the completed gene for HGH could be inserted.

To ready the synthetic fragment for combination with the RNA transcript, 10 μg of pHGH3 was cleaved with Eco RI and Hae III restriction endonucleases and the 77 base pair fragment containing coding sequences for HGH amino acids 1-23 was isolated from an 8% polyacrylamide gel.

The plasmid pHGH 31 (5 μg) was next cleaved with Hae III. The 551 bp HGH sequence and a comigrating 540 bp Hae III fragment of pBR322 were purified by gel electrophoresis. Subsequent treatment with Xma I cleaved only the HGH sequence, removing 39 base pairs from the 3' noncoding region. The resulting 512 bp

fragment was separated from the 540 bp pBR322 *Hae* III piece by electrophoresis on a 6% polyacrylamide gel. 0.3 µg of the 77 bp *Eco* RI-*Hae* III fragment was polymerized with T4 ligase in a 16 µl reaction vessel for 14 hours at 4° C. The mixture was heated to 70° C. for 5' to inactivate the ligase, then treated with *Eco* RI (to cleave fragments which had dimerized through their *Eco* RI sites) and with *Sma* I (to cleave *Xma* I dimers), yielding a 591 bp fragment with an *Eco* RI "cohesive" end and a *Sma* I "blunt" end. After purification on a 6% polyacrylamide gel, approximately 30 ng of this fragment were obtained. It should be noted that the expression plasmid pGH6 contains no *Xma* I recognition site. However, *Sma* I recognizes the same site as *Xma* I, but cuts through the middle of it, yielding blunt ends. The *Sma*-cleaved terminus of the fragment derived from pGH31 can accordingly be blunt end ligated into pGH6.

The expression plasmid pGH6, containing tandem lac UV5 promoters, was treated successively with *Hind* III, nuclease S1, and *Eco* RI and purified by gel electrophoresis. 50 ng of the resulting vector, which had one *Eco* RI cohesive end and one blunt end was ligated to 10 ng of the 591 bp HGH DNA. The ligation mixture was used to transform *E. coli* ×1776. Colonies were selected for growth on tetracycline (12.5 µg/ml). It is noteworthy that insertion of the hybrid HGH gene into pGH6 destroys the promoter for the tetracycline resistance gene, but that the tandem lac promoter permits read-through of the structural gene for tet resistance, retaining this selection characteristic. Approximately 400 transformants were obtained. Filter hybridization by the Grunstein-Hogness procedure, *Proc. Nat'l. Acad. Sci. USA*, 72 3961 (1975) identified 12 colonies containing HGH sequences. The plasmids isolated from three of these colonies gave the expected restriction patterns when cleaved with *Hae* III, *Pvu* II, and *Pst* I. The DNA sequence of one clone, pGH107, was determined.

Human growth hormone expressed by the transformants was easily detected by direct radioimmunoassay performed on serial dilutions of lysed cell supernatants using the Phadebas HGH PRIST kit (Pharmacia).

To demonstrate that HGH expression is under the control of the lac promoter, pGH107 was transformed into *E. coli* strain D1210 a lac⁺(i00+z+y⁺), a lac repressor overproducer. Meaningful levels of HGH expression could not be detected until addition of the inducer IPTG (isopropylthiogalactoside).

Removal of the *Eco* RI site in pGH107 would leave the ATG start signal the same distance from the ribosome binding site condons of the lac promoter as occurs in nature between those condons and the start signal for B-galactosidase. To determine whether expression would be increased by mimicking this natural spacing we converted pGH107 to pGH107-1 by opening the former with *Eco* RI, digesting the resulting single strand ends with S1 endonuclease, and recircularizing by blunt-end ligation with T4 ligase. Although the resulting plasmid proved likewise capable of expressing HGH, it surprisingly did so to a lesser extent than did pGH107, as shown by direct radioimmunoassay.

It will be apparent to those skilled in the art that the present invention is not limited to the preferred embodiment just discussed, but rather only to the lawful scope of the appended claims. Variations other than those hitherto discussed will be apparent, whether in the choice of promoter system, parental plasmid, intended polypeptide product or elsewhere. For example, other

promoter systems applicable to the present invention include the lambda promoter, the arabinose operon (phi 80 d ara) or the colicin E1, galactose, alkaline phosphatase or tryptophan promoter systems. Host organisms for bacterial expression may be chosen, e.g., from among the Enterobacteriaceae, such as strains of *Escherichia coli* and *Salmonella*; *Bacillaceae*, such as *Bacillus subtilis*; *Pneumococcus*; *Streptococcus*; and *Haemophilus influenzae*. Of course, the choice of organism will control the levels of physical containment in cloning and expression that should be practiced to comply with National Institutes of Health Guidelines for Recombinant DNA, 43 Fed. Reg. 60,080 (1978).

While preferred for bench-scale practice of the present invention, *E. coli* ×1776 could prove of limited practicality in large-scale industrial manufacture owing to the debilitations purposefully incorporated in it for biosafety reasons. With appropriate levels of physical, rather than biological, containment such organisms as *E. coli* K-12 strain 294, supra, and *E. coli* strain RR1, genotype: Pro⁻Leu⁻Thi⁻R_B⁻recA⁺Str^r Lac y⁻ could be employed in larger scale operation. *E. coli* RR1 is derived from *E. coli* HB101 (H. W. Boyer, et al., *J. Mol. Bio.* (1969) 41 459-472) by mating with *E. coli* K12 strain KL16 as the Hfr donor. See J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor, N.Y., 1972). A culture of *E. coli* RR1 was deposited Oct. 30, 1978 with the American Type Culture Collection, without restriction as to access (ATCC No. 31343). A culture of ×1776 was similarly deposited July 3, 1979 in the American Type Culture Collection (ATCC No. 31537). Deposits of the following were made in the American Type Culture Collection July 3, 1979: plasmid pGH107 (ATCC No. 40011); plasmid pGH6 (ATCC No. 40012); strain ×1776 transformed with pGH107 (ATCC No. 31538) and *E. coli* K12 strain 294 transformed with pGH6 (ATCC No. 31539).

Organisms produced according to the invention may be employed in industrial scale fermentative production of human growth hormone, yielding product in quantities and for applications hitherto unattainable. For example, transformant *E. coli* cultures may be grown up in aqueous media in a steel or other fermentation vessel conventionally aerated and agitated, in aqueous media at, e.g., about 37° C. and near neutral pH (e.g., pH 7±0.3) supplied with appropriate nutriment such as carbohydrate or glycerol, nitrogen sources such as ammonium sulfate, potassium sources such as potassium phosphate, trace elements, magnesium sulfate and the like. Transformant organisms preferably exhibit one or more selection characteristics, such as antibiotic resistance, so that selection pressures may be imposed to discourage competitive growth of wild-type *E. coli*. As an example, in the case of an ampicillin or tetracycline-resistant organism the antibiotic may be added to the fermentation medium to select out wild-type organisms which lack the resistance characteristic.

Upon completion of fermentation the bacterial suspension is centrifuged or the cellular solids otherwise collected from the broth and then lysed by physical or chemical means. Cellular debris is removed from supernatant and soluble growth hormone isolated and purified.

Human growth hormone may be purified from bacterial extracts using one or a combination of (1) polyethyleneimine fractionation; (2) gel filtration chromatography on Sephacryl S-200; (3) ion exchange chromatography on Biorex-70 resin or CM Sephadex; (4) ammonium

sulphate and/or pH fractionation; and (5) affinity chromatography using antibody resins prepared from anti-HGH IgG isolated from immunosensitized animals or hybridomas; and desorbed under acid or slightly denaturing conditions.

We claim:

1. In the method of constructing a replicable cloning vehicle capable, in a microbial organism, of expressing a particular polypeptide of known amino acid sequence wherein a gene coding for the polypeptide is inserted into a cloning vehicle and placed under the control of an expression promoter, the improvement which comprises:

- (a) obtaining by reverse transcription from messenger RNA a first gene fragment for an expression product other than said polypeptide, which fragment comprises at least a portion of the coding sequence for said polypeptide;
- (b) where the first fragment comprises protein-encoding codons for amino acid sequences other than those contained in said polypeptide, eliminating the same while retaining at least a substantial portion of said coding sequence, the resulting fragment nevertheless coding for an expression product other than said polypeptide;
- (c) providing by organic synthesis one or more synthetic non-reverse transcript-gene fragments encoding the remainder of the amino acid sequence of said polypeptide, at least one of said fragments coding for the amino-terminal portion of the polypeptide; and
- (d) deploying the synthetic gene fragment(s) of step (c) and that produced in step (a) or (b), as the case may be, in a replicable cloning vehicle in proper reading phase relative to one another and under the control of an expression promoter;

whereby a replicable cloning vehicle capable of expressing the amino acid sequence of said polypeptide is formed.

2. The method of claim 1 wherein the cloning vehicle of step (d) is a bacterial plasmid.

3. The method of claim 2 wherein the synthetic fragment encoding the amino-terminal portion of the poly-

peptide additionally codes for expression of a specifically cleavable amino acid sequence, and wherein the fragments are deployed downstream from and in reading phase with expressed protein-encoding condons, whereby the conjugated plasmid expression product may be specifically cleaved to yield the polypeptide.

4. The method of claim 2 wherein the amino acid sequence of the polypeptide is expressible unaccompanied by extraneous protein.

5. The method of claim 4 wherein the fragment of step (a) comprises at least a majority of the coding sequence for said polypeptide.

6. The method of claim 2 wherein a synthetic fragment and an mRNA transcript fragment are ligated to one another before their deployment in the cloning vehicle, and wherein the opposite ends of the fragment and of the transcript are variously single stranded or blunt so as to ensure ligation of the two fragments in the proper order for expression of said polypeptide.

7. The method of claim 5 wherein the polypeptide is human growth hormone, and wherein the first fragment comprises protein-encoding codons for amino acid sequences other than those in human growth hormone, and wherein elimination step (b) yields the Hae III restriction enzyme fragment of the first fragment.

8. The method of claim 7 wherein step (b) includes digestion of the Hae III fragment with a different restriction enzyme, cleaving away codons for untranslated messenger RNA and simultaneously providing a single-stranded terminus at one end of the resulting fragment.

9. The method of claim 8 wherein the second restriction enzyme is Xma I.

10. A method according to claim 1 wherein the polypeptide is human growth hormone and wherein the codons for amino acids 1-24 thereof are essentially as depicted in FIG. 1.

11. A method according to claim 4 wherein the polypeptide is human growth hormone and wherein the codons for amino acids 1-24 thereof are essentially as depicted in FIG. 1.

12. A method according to claim 7 wherein the codons for amino acids 1-24 are essentially as depicted in FIG. 1.

* * * * *

TAB E E

- (21) Application No 7843446
 (22) Date of filing 6 Nov 1978
 (23) Claims filed 6 Nov 1978
 (30) Priority data
 (31) 849692
 (32) 8 Nov 1977
 (33) United States of America (US)
 (43) Application published 23 May 1979
 (51) INT CL²
 C07H 21/00 C07C 103/52
 C07G 7/00 7/02 15/00
 (52) Domestic classification C3H B
 (56) Documents cited
 GB 1521032
 Science 198 1056—1062 (1977)
 Pro-Nat Acad Sci., USA 71 3455 (1974)
 Scientific American 233,24(1975)
 Nature, 270 494—499 (1977)
 (58) Field of search C3H

- (71) Applicants
 Genentech, Inc., 460
 Point San Bruno
 Boulevard, So. San
 Francisco, California,
 United States of America
 (72) Inventors
 Keiichi Itakura, Arthur
 Dale Riggs
 (74) Agents
 Abel & Imray

(54) Method for microbial polypeptide expression

(57) A recombinant plasmid suited for transformation of a bacterial host and

use therein as a cloning vehicle, wherein the plasmid comprises:

a) a regulon homologous to the bacterial host in its untransformed state; and

b) in reading phase with the regulon, a DNA insert coding for the amino acid sequence of a heterologous polypeptide, such that bacteria transformed by the plasmid are capable of expressing said amino acid sequence in recoverable form. The heterologous polypeptide is, for example, a mammalian hormone, or an intermediate therefor.

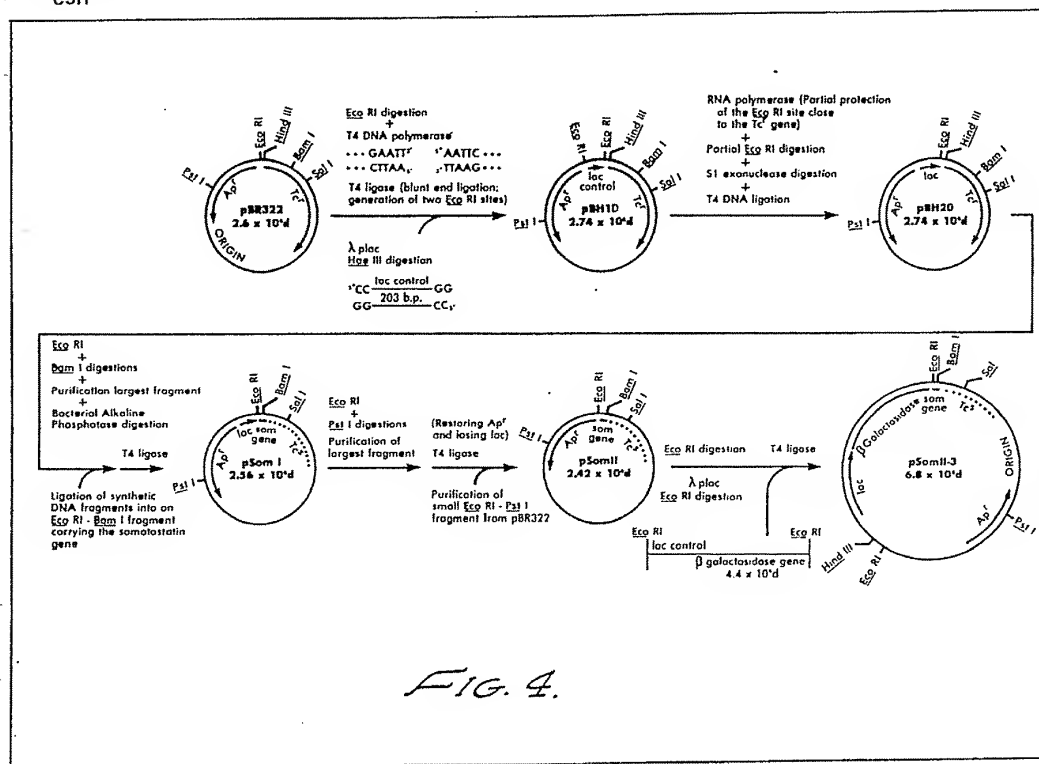


FIG. 4.

GB 2 007 676 A

(21) Application No 7843446
 (22) Date of filing 6 Nov 1978
 (23) Claims filed 6 Nov 1978
 (30) Priority data
 (31) 849692
 (32) 8 Nov 1977
 (33) United States of America (US)
 (43) Application published 23 May 1979
 (51) INT CL²
 C07H 21/00 C07C 103/52
 C07G 7/00 7/02 15/00
 (52) Domestic classification C3H B
 (56) Documents cited
 GB 1521032
 Science 198 1056—1062 (1977)
 Pro-Nat Acad Sci., USA 71 3455 (1974)
 Scientific American 233,24(1975)
 Nature, 270 494—499 (1977)
 (58) Field of search C3H

(71) Applicants
 Genentech, Inc., 460
 Point San Bruno
 Boulevard, So. San
 Francisco, California,
 United States of America
 (72) Inventors
 Keiichi Itakura, Arthur
 Dale Riggs
 (74) Agents
 Abel & Inray

(54) Method for microbial polypeptide expression

(57) A recombinant plasmid suited for transformation of a bacterial host and

use therein as a cloning vehicle, wherein the plasmid comprises:

a) a regulon homologous to the bacterial host in its untransformed state; and

b) in reading phase with the regulon, a DNA insert coding for the amino acid sequence of a heterologous polypeptide, such that bacteria transformed by the plasmid are capable of expressing said amino acid sequence in recoverable form. The heterologous polypeptide is, for example, a mammalian hormone, for example, somatostatin, or an intermediate therefor.

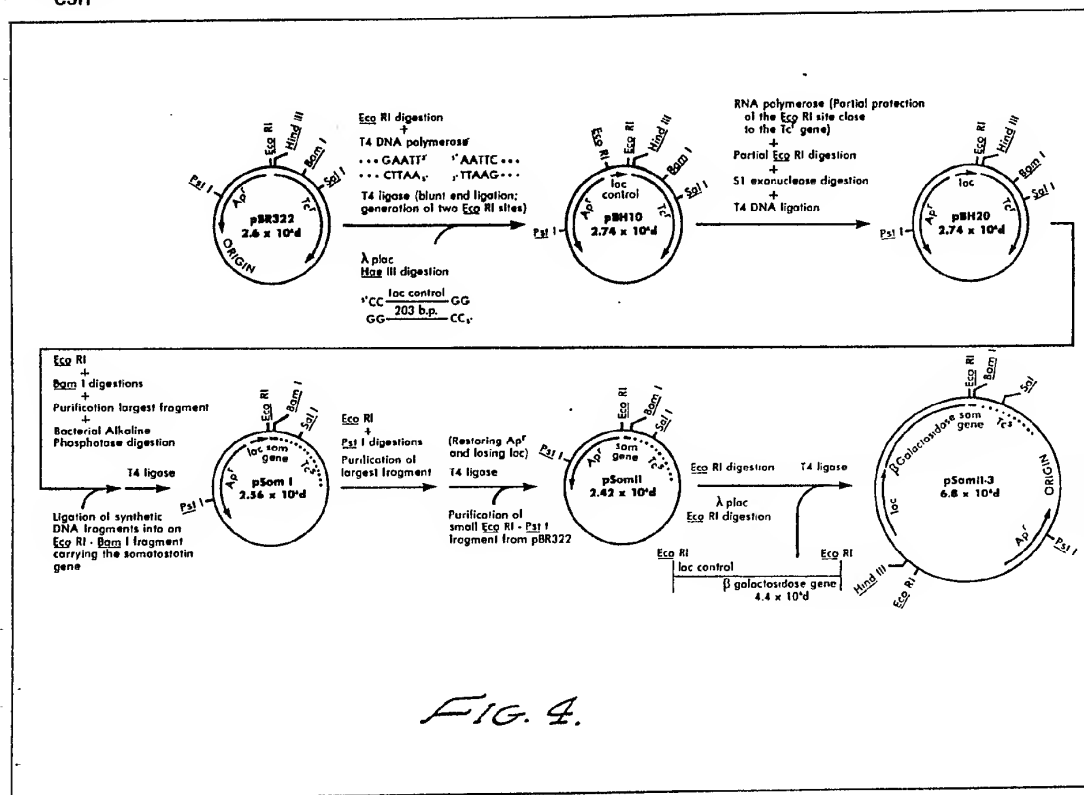


FIG. 4.

GB 2 007 676 A



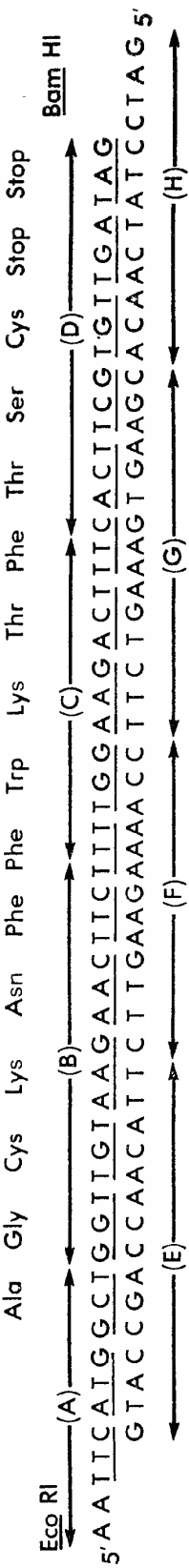


FIG. 2.

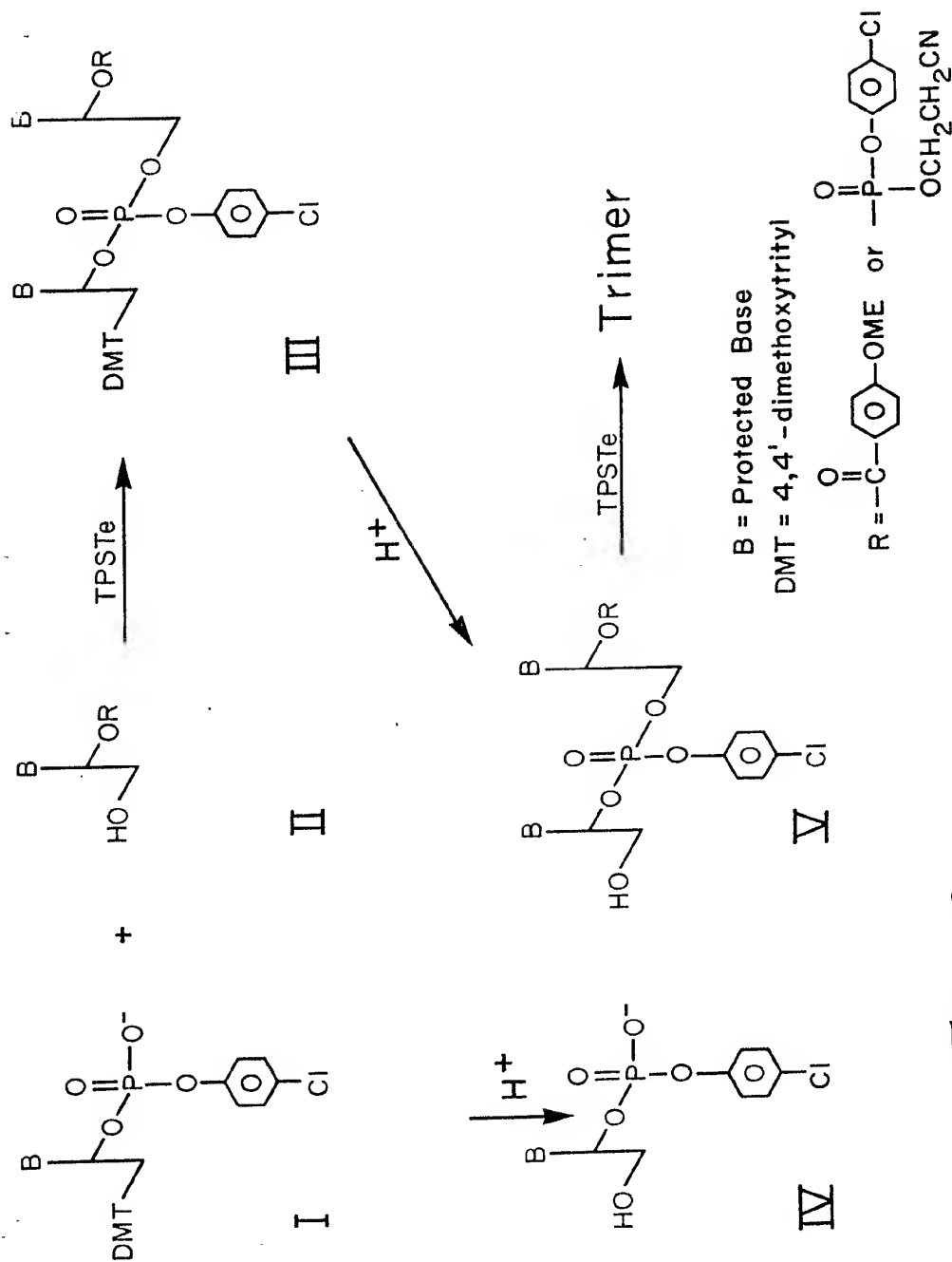


FIG. 3.

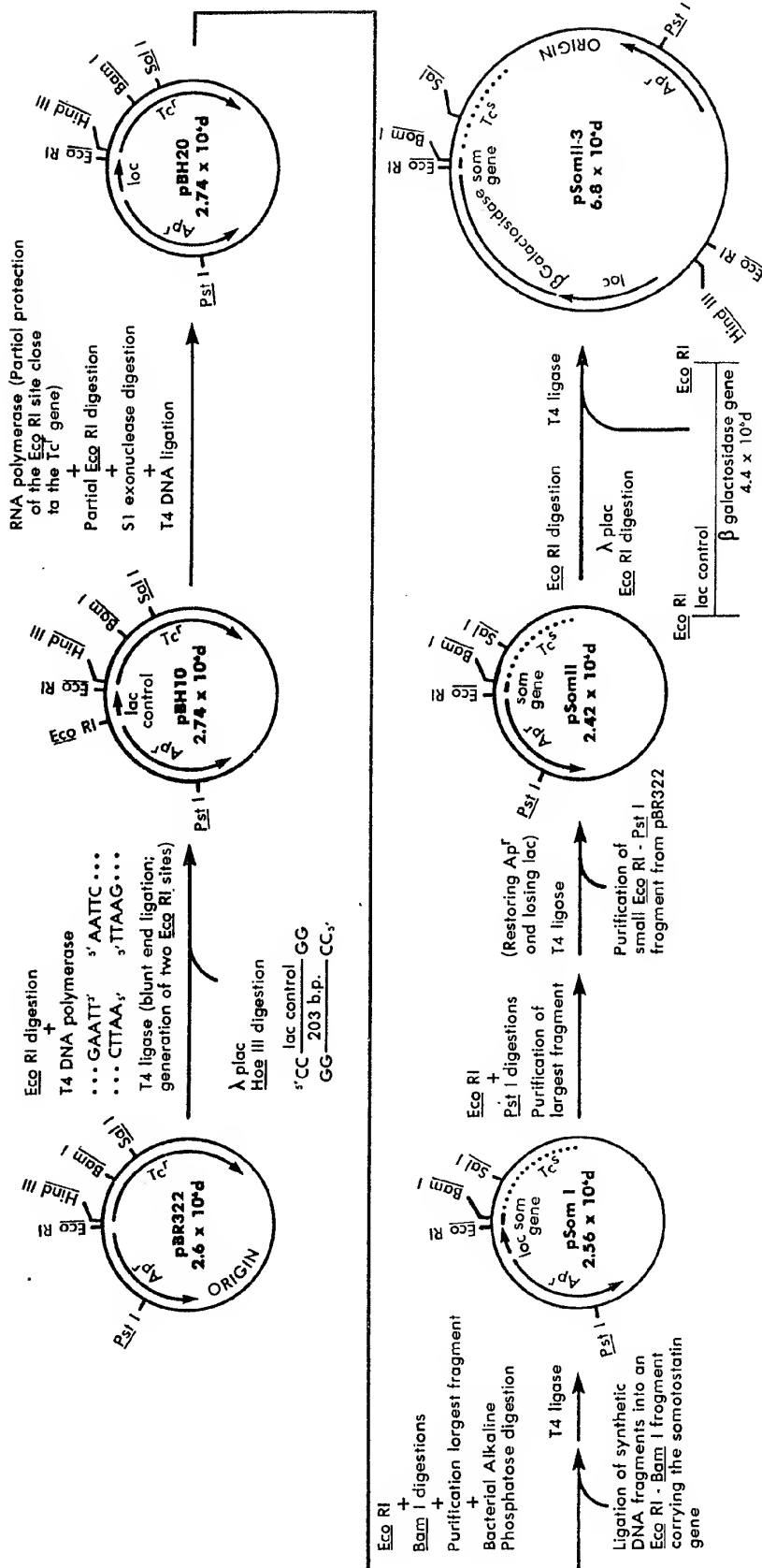


Fig. 4.

FIG. 5A.

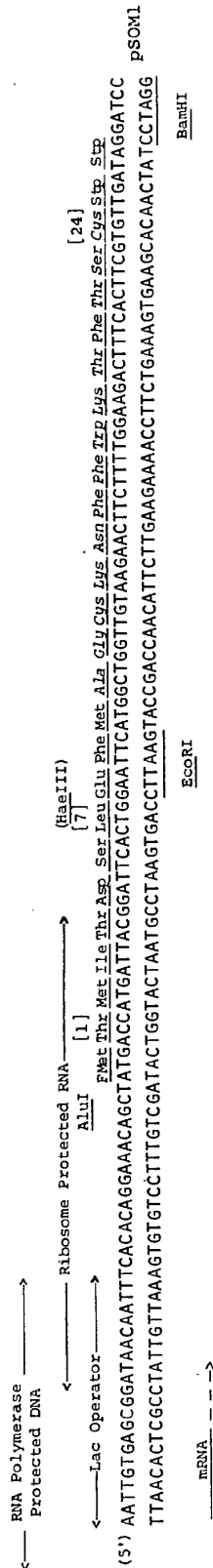


FIG. 5B.

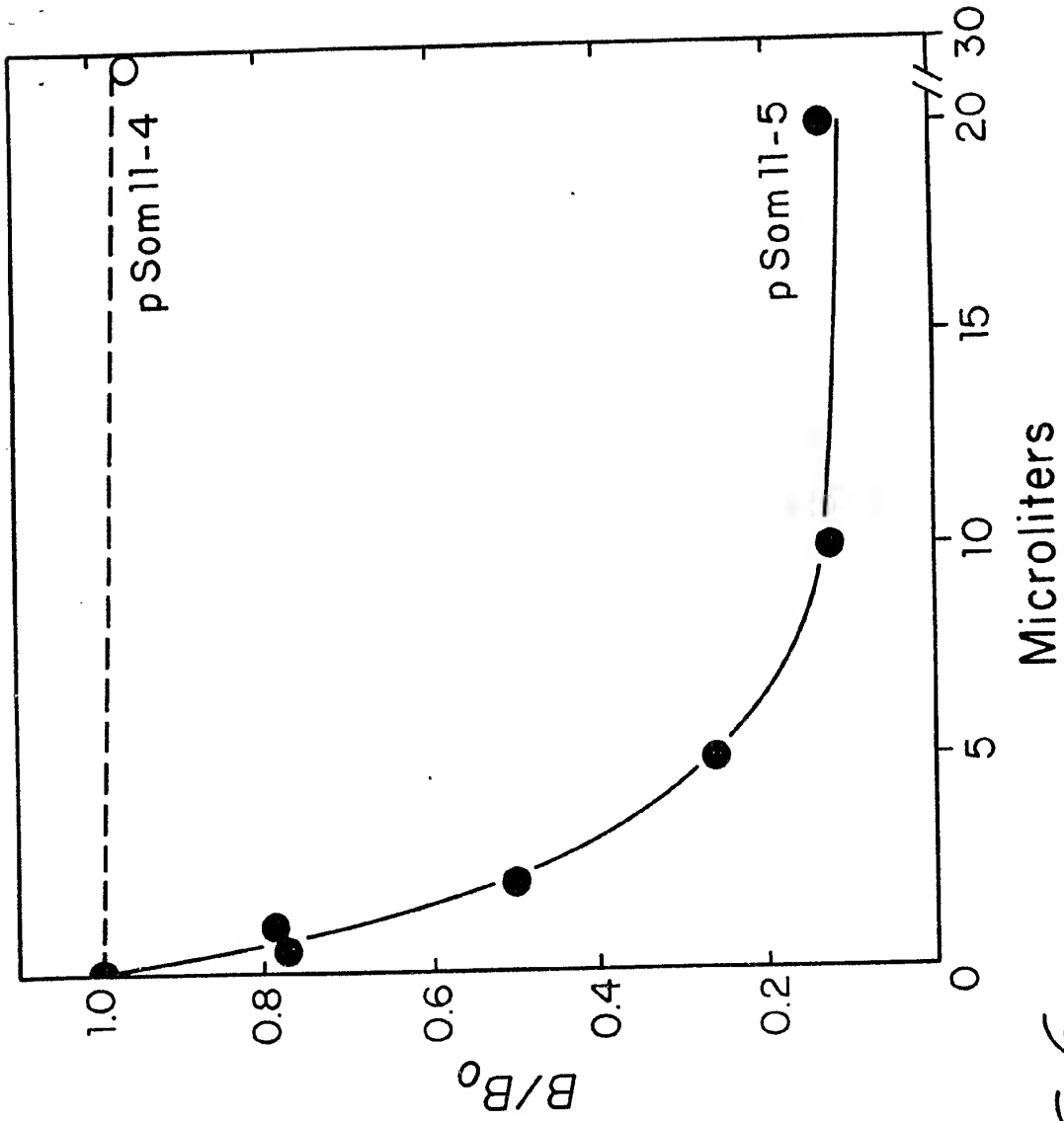


FIG. 6.

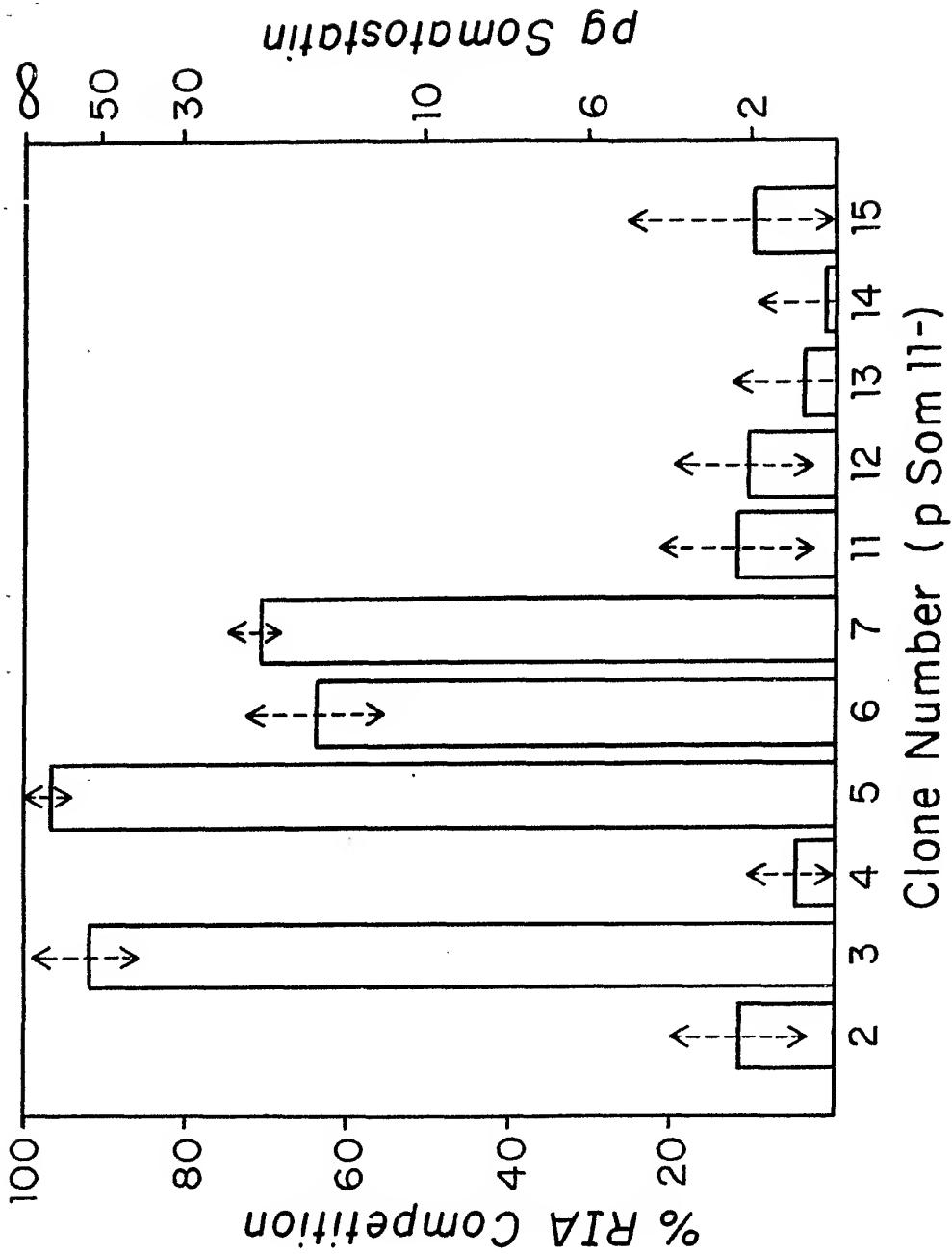


Fig. 7

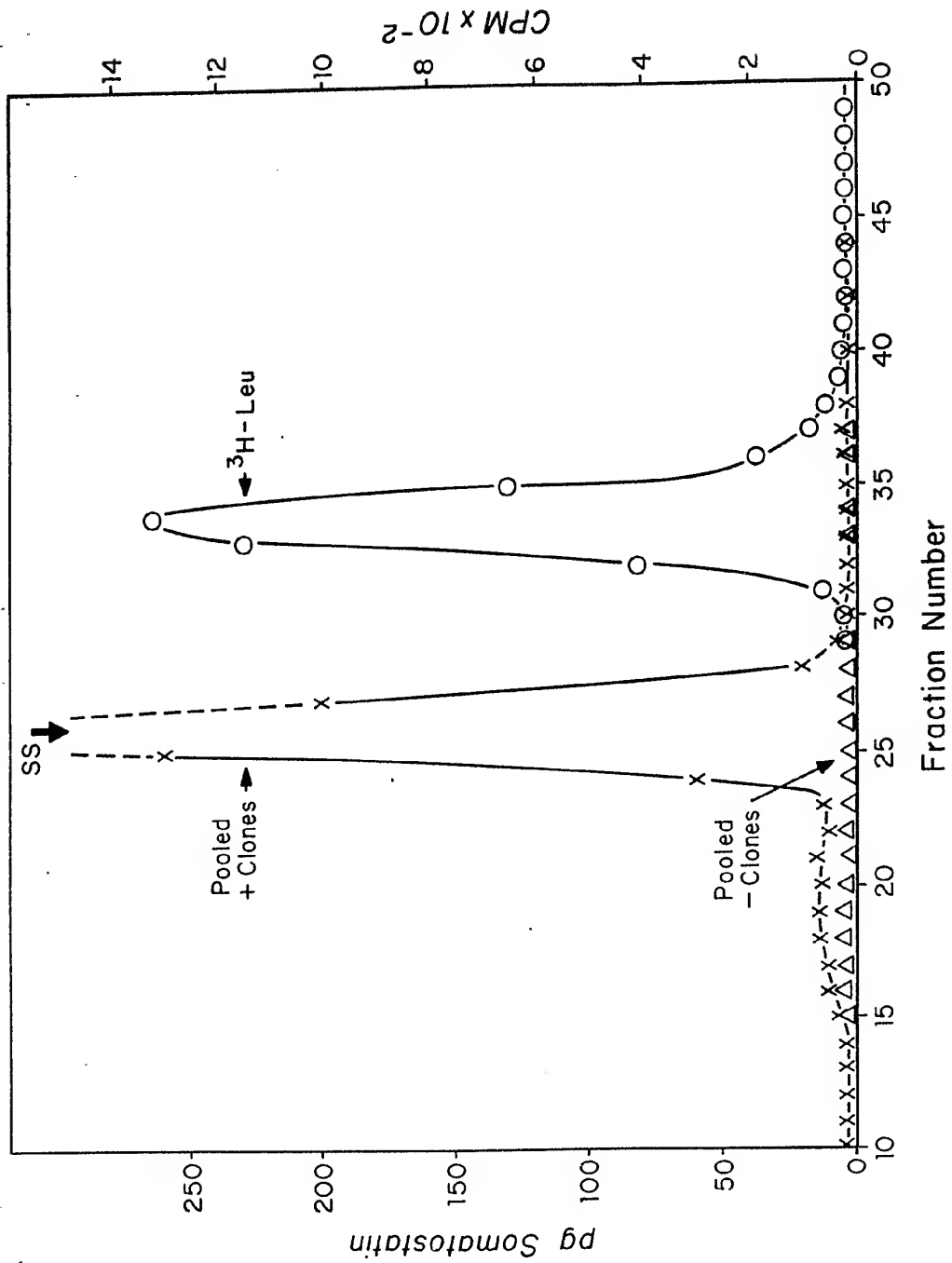


Fig. 8.

A-Chain Gene

III. PÜH

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
 Met Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
 Eco RI
 A₁ A₂ A₃ A₄ A₅ A₆ stop stop
 AATTCATGGCATCGTTGAACAGTGTTGCACCTTCTATCTGCTCTCTTACCAGCTTGAGAACACTGTAACATAATAG
 GTACCCGTAGCAACTTGTACAACTGAAGATAGCAGGAGAAATGTCGAACTCTTGATGACATTGATTATCTCTAG
 A₇ A₈ A₉ A₁₀ A₁₁ A₁₂ Bam I

Fig. 9.

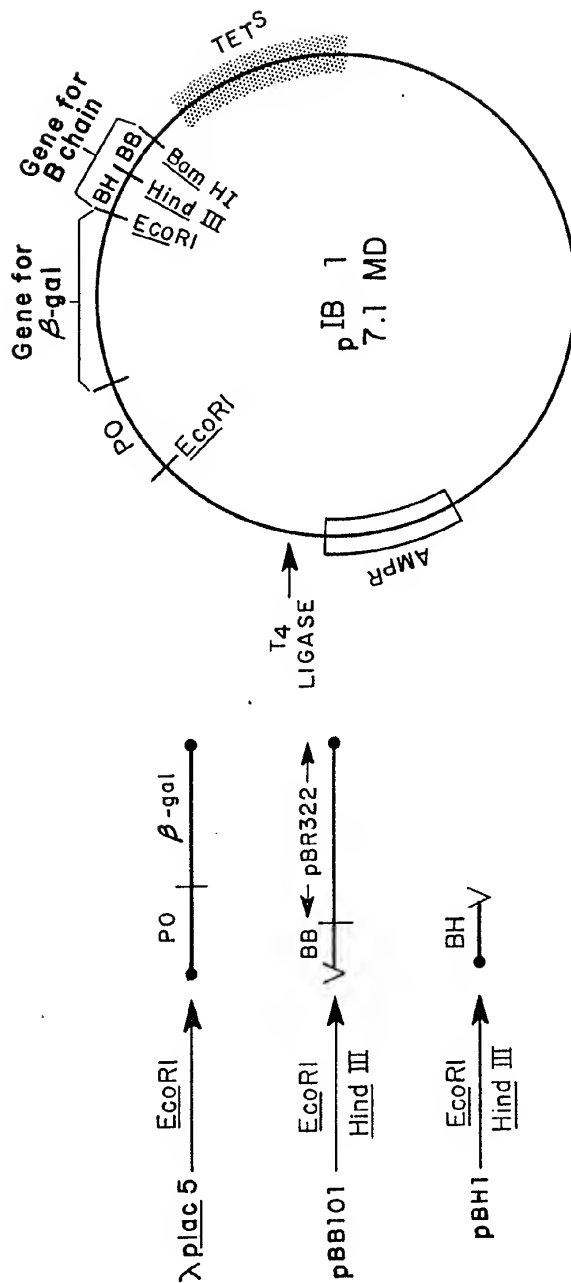


FIG. 10.

SPECIFICATION

Method and means for microbial polypeptide expression

This invention relates to a method and means for microbial polypeptide expression.

Genetic information is encoded on double-stranded deoxyribonucleic acid ("DNA" or "genes") according to the order in which the DNA coding strand presents the characteristic bases of its repeating nucleotide components. "Expression" of the encoded information to form polypeptides involves a two-part process. According to the dictates of certain control regions ("regulons") in the gene, RNA polymerase may be caused to move along the coding strand, forming messenger RNA (ribonucleic acid) in a process called "transcription". In a subsequent "translation" step the cell's ribosomes in conjunction with transfer RNA convert the mRNA "message" into polypeptides. Included in the information mRNA transcribes from DNA are signals for the start and termination of ribosomal translation, as well as the identity and sequence of the amino acids which make up the polypeptide. The DNA coding strand comprises long sequences of nucleotide triplets called "codons" because the characteristic bases of the nucleotides in each triplet or codon encode specific bits of information. For example, 3 nucleotides read as ATG (adenine-thymine-guanine) result in an mRNA signal interpreted as "start translation", while termination codons TAG, TAA and TGA are interpreted "stop translation". Between the start and stop codons lie the so-called structural gene, whose codons define the amino acid sequence ultimately translated. That definition proceeds according to the well-established "genetic code" (e.g., J. D. Watson, *Molecular Biology of the Gene* W. A. Benjamin Inc., N. Y., 3rd ed. 1976) which describes the codons for the various amino acids. The genetic code is degenerate in the sense that different codons may yield the same amino acid, but precise in that for each amino acid there are one or more codons for it and no other. Thus, for example, all of the codons TTT, TTC, TTA and TTG, when read as such, encode for serine and no other amino acid. During transcription the proper reading phase or reading frame must be maintained. Consider for example what happens when the RNA polymerase transcriber reads different bases as the beginning of a codon (underlined) in the sequence ...

```

... GCT GGT TGT AAG ... — ... Ala-Gly-Cys-Lys ...
... G CTG GTT GTA AG ... — ... Leu-Val-Leu ...
... GC TGG TTG TAA A ... — ... Trp-Leu-(STOP) ...

```

The polypeptide ultimately produced, then, depends vitally upon the spatial relationship of the structural gene with respect to the regulon.

A clearer understanding of the process of genetic expression will emerge once certain components of genes are defined:

Operon — A gene comprising structural gene(s) for polypeptide expression and the control region ("regulon") which regulates that expression.

Promoter — A gene within the regulon to which RNA polymerase must bind for initiation of transcription.

Operator — A gene to which repressor protein may bind, thus preventing RNA polymerase binding on the adjacent promoter.

Inducer — A substance which deactivates repressor protein, freeing the operator and permitting RNA polymerase to bind to promoter and commence transcription.

Catabolite Activator Protein ("CAP") Binding Site — A gene which binds cyclic adenosine monophosphate ("c AMP")—mediated CAP, also commonly required for initiation of transcription. The CAP binding site may in particular cases be unnecessary. For example, a promoter mutation in the lactose operon of the phage λ plac UV5 eliminates the requirement for cAMP and CAP expression. J. Beckwith et al, *J. Mol. Biol* 69, ISS—160 (1972).

Promoter-Operator System — As used herein, an operable control region of an operon, with or without respect to its inclusion of a CAP binding site or capacity to code for repressor protein expression.

Further by way of definition, and for use in the discussion of recombinant DNA which follows, we define the following:

Cloning Vehicle — Non-chromosomal double stranded DNA comprising an intact "replicon" such that the vehicle is replicated, when placed within a unicellular organism ("microbe") by a process of "transformation". An organism so transformed is called a "transformant".

Plasmid — For present purposes, a cloning vehicle derived from viruses or bacteria, the latter being "bacterial plasmids."

Complementarity — A property conferred by the base sequences of single strand DNA which permits the formation of double stranded DNA through hydrogen bonding between complementary bases on the respective strands. Adenine (A) complements thymine (T), while guanine (G) complements cytosine (C).

Advances in biochemistry in recent years have led to the construction of "recombinant" cloning vehicles in which, for example, plasmids are made to contain exogenous DNA. In particular instances the recombinant may include "heterologous" DNA, by which is meant DNA that codes for polypeptides

ordinarily not produced by the organism susceptible to transformation by the recombinant vehicle. Thus, plasmids are cleaved to provide linear DNA having ligatable termini. These are bound to an exogenous gene having ligatable termini to provide a biologically functional moiety with an intact replicon and a desired phenotypical property. The recombinant moiety is inserted into a microorganism by

5 transformation and transformants are isolated and cloned, with the object of obtaining large populations 5 capable of expressing the new genetic information. Methods and means of forming recombinant cloning vehicles and transforming organisms with them have been widely reported in the literature. See, e.g., H. L. Heynecker et al, *Nature* 263, 748—752 (1976); Cohen et al, *Proc. Nat. Acad. Sci. USA* 69, 2110 (1972); *ibid.*, 70, 1293 (1973); *ibid.*, 70, 3240 (1973); *ibid.*, 71, 1030 (1974); Morrow et al, *Proc. Nat. Acad. Sci. U.S.A.* 71, 1743 (1974); Novick, *Bacteriological Rev.*, 33, 210 (1969); Hersfield et al, *Proc. Soc. Nat'l. Acad. Sci. U.S.A.* 71, 3455 (1974) and Jackson et al, *ibid.* 69, 2904 (1972). A generalised

10 discussion of the subject appears in S. Cohen, *Scientific American* 233, 24 (1975). These and other 10 publications alluded to herein are incorporated by reference.

A variety of techniques are available for DNA recombination, according to which adjoining ends of

15 separate DNA fragments are tailored in one way or another to facilitate ligation. The latter term refers to 15 the formation of phosphodiester bonds between adjoining nucleotides, most often through the agency of the enzyme T4 DNA ligase. Thus, blunt ends may be directly ligated. Alternatively, fragments containing complementary single strands at their adjoining ends are advantaged by hydrogen bonding

20 which positions the respective ends for subsequent ligation. Such single strands, referred to as cohesive 20 termini, may be formed by the addition of nucleotides to blunt ends using terminal transferase, and sometimes simply by chewing back one strand of a blunt end with an enzyme such as λ -exonuclease. Again, and most commonly, resort may be had to restriction endonucleases, which cleave

25 phosphodiester bonds in and around unique sequences of nucleotides of about 4—6 base pairs in 25 length. Many restriction endonucleases and their recognition sites are known, the so-called *Eco* RI endonucleases being most widely employed. Restriction endonucleases which cleave double-stranded DNA at rotationally symmetric "palindromes" leave cohesive termini. Thus, a plasmid or other cloning

30 vehicle may be cleaved, leaving termini each comprising half the restriction endonuclease recognition 30 site. A cleavage product of exogenous DNA obtained with the same restriction endonuclease will have ends complementary to those of the plasmid termini. Alternatively, as disclosed *infra*, synthetic DNA comprising cohesive termini may be provided for insertion into the cleaved vehicle. To discourage

35 rejoiner of the vehicles' cohesive termini pending insertion of exogenous DNA, the termini can be 35 digested with alkaline phosphatase, providing molecular selection for closures incorporating the exogenous fragment. Incorporation of a fragment having the proper orientation relative to other aspects of the vehicle may be enhanced when the fragment supplants vehicle DNA excised by two different

restriction endonucleases, and itself comprises termini respectively constituting half the recognition

sequence of the different endonucleases.

Despite wide-ranging work in recent years in recombinant DNA research, few results susceptible to immediate and practical application have emerged. This has proven especially so in the case of failed attempts to express polypeptides and the like coded for by "synthetic DNA", whether constructed

40 nucleotide by nucleotide in the conventional fashion or obtained by reverse transcription from isolated 40 *in* RNA (complementary or "cDNA"). In this application we describe what appears to represent the first expression of a functional polypeptide product from a synthetic gene, together with related developments which promise wide-spread application. The product referred to is somatostatin (Guillemin U.S.P. 3,904,594), and inhibitor of the secretion of growth hormone, insulin and glucagon

45 whose effects suggest its application in the treatment of acromelagy, acute pancreatitis and insulin- 45 dependent diabetes. See R. Guillemin et al, *Annual Rev. Med.* 27 379 (1976). The somatostatin model clearly demonstrates the applicability of the new developments described here on numerous and beneficial fronts, as will appear from the accompanying drawings and more clearly from the detailed description which follows.

50 Summary of Invention 50

According to the invention there is provided a recombinant plasmid suited for transformation of a bacterial host and use therein as a cloning vehicle, wherein the plasmid comprises:

- a) A regulon homologous to the bacterial host in its untransformed state; and
 - b) In reading phase with the regulon, a DNA insert coding for the amino acid sequence of a
- 55 heterologous polypeptide, such that bacteria transformed by the plasmid are capable of expressing said 55 amino acid sequence in recoverable form.

Brief Description of the Drawings

The accompanying drawings illustrate one context in which preferred embodiments of the invention find application, i.e., expression of the hormone somatostatin by bacterial transformants.

60 containing recombinant plasmids. 60

Figure 1. Schematic outline of the process: the gene for somatostatin, made by chemical DNA synthesis, is fused to the *E. coli* β -galactosidase gene on the plasmid pBR322. After transformation into *E. coli*, the recombinant plasmid directs the synthesis of a precursor protein which can be specifically

cleaved *in vitro* at methionine residues by cyanogen bromide to yield active mammalian polypeptide hormone. A, T, C and G denote the characteristic bases (respectively adenine, thymine, cytosine and guanine) of the deoxyribonucleotides in the coding strand of the somatostatin gene.

Figure 2. Schematic structure of a synthetic gene whose coding strand (i.e., the "upper" strand) comprises codons for the amino acid sequence of somatostatin (given).

Figure 3. Schematic illustration of preferred method for construction of nucleotide trimers used in constructing synthetic genes. In the conventional notation employed to depict nucleotides in Figure 3, the 5' OH is to the left and the 3' OH to the right, e.g.

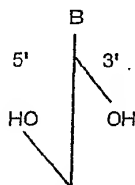


Figure 4. Flow chart for the construction of a recombinant plasmid (e.g., pSOM11—3) capable of expressing a somatostatin ("SOM")-containing protein, beginning with the parental plasmid pBR322. In Figure 4 the approximate molecular weight of each plasmid is stated in daltons ("d"). Ap^r and Tc^r respectively denote genes for ampicillin and tetracycline resistance, while Tc^s denotes tetracycline susceptibility resulting from excision of a portion of the Tc^r gene. The relative positions of various restriction endonuclease specific cleavage sites on the plasmids are depicted (e.g., *Eco* RI, *Bam* I, etc.).

Figures 5A and 5B. The nucleotide sequences of key portions of two plasmids are depicted, as is the direction of messenger RNA ("mRNA") transcription, which invariably proceeds from the 5' end of the coding strand. Restriction endonuclease substrate sites are as shown. Each depicted sequence contains both the control elements of the lac (lactose) operon, and codons for expression of the amino acid sequence of somatostatin (italics). The amino acid sequence numbers for β -galactosidase (" β -gal") are in brackets.

Figures 6—8. As more particularly described in the "Experimental" discussion, *infra*, these depict the results of comparative radioimmune assay experiments which demonstrate the somatostatin activity of product expressed by the recombinant plasmids.

Figure 9. Schematic structure of synthetic genes whose coding strands comprise codons for the amino acid sequences of the A and B strands of human insulin.

Figure 10. Flow chart for construction of a recombinant plasmid capable of expressing the B chain of human insulin.

Detailed Description

30 1. Preparation of Genes Coding for Heterologous Polypeptide 30

DNA coding for any polypeptide of known amino acid sequence may be prepared by choosing codons according to the genetic code. For ease in purification, etc., oligodeoxyribonucleotide fragments of, for example, from about 11 to about 16 nucleotides are prepared separately, then assembled in the desired sequence. Thus, one prepares first and second series of oligodeoxyribonucleotide fragments of convenient size. The first series, when joined in proper sequence, yield a DNA coding strand for polypeptide expression (see, e.g., Figure 2, fragments A, B, C and D). The second series, when likewise joined in proper sequence, yield a strand complementary to the coding strand (e.g., Figure 2, fragments E, F, G and H). The fragments of the respective strands preferably overlap such that complementarity promotes their self assembly through hydrogen bonding of the cohesive termini of fragment blocks.

40 Following assembly, the structural gene is completed by ligation in the conventional manner. 40

The degeneracy of the genetic code permits substantial freedom in the choice of codons for any given amino acid sequence. For present purposes, however, codon choice was advantageously guided by three considerations. First, codons and fragments were selected, and fragment assembly was staged, so as to avoid undue complementarity of the fragments, one with another, save for fragments adjacent one another in the intended gene. Secondly, sequences rich in AT base pairs (e.g., about five or more) are avoided, particularly when produced by a sequence rich in GC base pairs, to avoid premature termination of transcription. Thirdly, at least a majority of the codons chosen are those preferred in the expression of microbial genomes (see, e.g., W. Fiers, et al, *Nature* 260, 500 (1976). For purposes of the appended claims, we define the following as codons "preferred for the expression of microbial genomes":

TABLE I
PREFERRED ASSIGNMENT OF CODONS

| First Position (5' End) (Read Down) | Second Position (Read Across) | | | | Third Position (3' End) (Read Down) |
|---|----------------------------------|--------------------------|--------------------------|-------------------------|---|
| | T | C | A | G | |
| T | phe phe leu — | — ser — ser | — tyr Stop Stop | cys — Stop trp | T C A G |
| C | leu. leu leu — | pro pro pro pro | his his gln gln | arg arg — — | T C A G |
| A | ile ile — met (start) | thr thr — thr | asn asn — lys | — ser — — | T C A G |
| G | val val val val | ala — — ala | asp asp glu glu | gly — — — | T C A G |

Most preferably in the case of somatostatin, the amino acid (codon) relationships of the structural gene are: gly (GGT); cys (TGT); lys (AAG); trp (TGG); ala (GCT, GCG); asn (AAT, AAC); phe (TTC, TTT); thr (ACT, ACG); and ser (TCC, TCG).

5. Where the structural gene of a desired polypeptide is to be inserted in a cloning vehicle for expression as such, the gene is preceded by a "start" codon (e.g., ATG) and immediately followed by one or more termination or stop codons (see Fig. 2). However, as described *infra*, the amino acid sequence of a particular polypeptide may be expressed with additional protein preceding and/or following it. If the intended use of the polypeptide requires cleavage of the additional protein, appropriate cleavage sites are coded for adjacent the polypeptide-additional protein codon junction. Thus, in Figure 1 as an example, the expression product is a precursor protein comprising both somatostatin and the greatest part of the β -galactosidase polypeptide. Here ATG is not required to code for the start of translation because ribosomal construction of the additional β -gal protein reads through into the somatostatin structural gene. Incorporation of the ATG signal, however, codes for the production of methionine, an amino acid specifically cleaved by cyanogen bromide, affording a facile method for converting precursor protein into the desired polypeptide.

- Figure 2 also exemplifies a further feature preferred in heterologous DNA intended for recombinant employment, i.e., the provision of cohesive termini, preferably comprising one of the two strands of a restriction endonuclease recognition site. For reasons previously discussed, the termini are preferably designed to create respectively different recognition sites upon recombination.

- While the developments described here have been demonstrated as successful with the somatostatin model, it will be appreciated that heterologous DNA coding for virtually any known amino acid sequence may be employed, *mutatis mutandis*. Thus, the techniques previously and hereafter discussed are applicable, *mutatis mutandis*, to the production of poly(amino)acids, such as polyleucine and polyalanine; enzymes; serum proteins; analgesic polypeptides, such as β -endorphins, which modulate thresholds of pain, etc. Most preferably, the polypeptides produced as such will be mammalian hormones or intermediates therefor. Among such hormones may be mentioned, e.g., somatostatin, human insulin, human and bovine growth hormone, leutinizing hormone, ACTH, pancreatic polypeptide, etc. Intermediates include, for example, human preproinsulin, human proinsulin, the A and B chains of human insulin and so on. In addition to DNA made *in vitro*, the heterologous DNA may comprise cDNA resulting from reverse transcription from mRNA. See, Ullrich et al, *Science* 196, 1313 (1977).

2. Recombinants Coding for the Expression of Precursor Protein

- In the process schematically depicted in Figure 1, expression yields a precursor protein comprising both a polypeptide coded for by a specific heterologous structural gene (somatostatin) and additional protein (comprising a portion of the β -galactosidase enzyme). A selective cleavage site adjacent the somatostatin amino acid sequence permits subsequent separation of the desired polypeptide from superfluous protein. The case illustrated is representative of a large class of procedures made available by the techniques described herein.

- Most commonly, cleavage will be effected outside the replicative environment of the Plasmid or

other vehicle as, for example, following harvest of the microbial culture. In this fashion temporary conjugation of small polypeptides with superfluous protein may preserve the former against, e.g., in vivo degradation by endogenous enzymes. At the same time, the additional protein will ordinarily rob the desired polypeptide of bioactivity pending extra-cellular cleavage, with the effect of enhancing the biosafety of the procedure. In particular instances, of course, it may prove desirable to effect cleavage within the cell. For example, cloning vehicles could be provided with DNA coding for enzymes which convert insulin precursors to the active form, operating in tandem with other DNA coded for expression of the precursor form.

In preferred case, the particular polypeptide desired lacks internal cleavage sites corresponding to that employed to shed superfluous protein, although it will be appreciated that where that condition is not satisfied competition reactions will yet give the desired product, albeit in lower yield. Where the desired product is methionine-free, cyanogen bromide cleavage at methionine adjacent the desired sequence has proven highly effective. Likewise, arginine- and lysine-free products may be enzymatically cleaved with, e.g., trypsin or chymotrypsin at arg-arg, lys-lys or like cleavage sites adjacent the desired sequence. In the case where cleavage leaves, e.g., unwanted arginin attached to desired product, it may be removed by carboxypeptidase digestion. When trypsin is employed to cleave an arg-arg, lysine sites within the desired polypeptide may first be protected, as with maleic or citraconic anhydrides. The cleavage techniques discussed here by way of example are but representative of the many variants which will occur to the art-skilled in light of the specification.

Cleavable protein may be expressed adjacent either the C- or N-terminals of a specific polypeptide, or even within the polypeptide itself, as in the case of the included sequence which distinguishes proinsulin and insulin. Again, the vehicle employed may code for expression of protein comprising repeated sequences of the desired polypeptide, each separated by selective cleavage sites. Most preferably, however, codons for superfluous protein will be translated in advance of the structural gene of the desired product, as in the case illustrated in the Figures. In every case care should be taken to maintain the proper reading frame relative to the regulon.

3. Expression of Immunogens

The ability to express both a specific polypeptide and superfluous protein provides useful tools for the production of immunogenic substances. Polypeptide "haptens" (i.e. substances containing determinants specifically bound by antibodies and the like but ordinarily too small to elicit an immune response) can be expressed as conjugates with additional protein sufficient in size to confer immunogenicity. Indeed, the β -gal-somatostatin conjugate produced here by way of example is of immunogenic size and may be expected to raise antibodies which bind the somatostatin hapten. Proteins comprising in excess of 100 amino acids, most commonly in excess of 200 such, exhibit immunogenic character.

Conjugates prepared in the foregoing fashion may be employed to raise antibodies useful in radioimmune or other assays for the hapten, and alternatively in the production of vaccines. We next describe an example of the latter application. Cyanogen bromide—or other cleavage products of viral protein will yield oligopeptides which bind to antibody raised to the protein itself. Given the amino acid sequence of such an oligopeptide hapten, heterologous DNA therefore may be expressed as a conjugate with additional protein which confers immunogenicity. Use of such conjugates as vaccines could be expected to diminish side reactions which accompany use of coat protein itself to confer immunity.

4. The Control Elements

Figure 1 depicts a process wherein a transformant organism expresses polypeptide product from heterologous DNA brought under the control of a regulon "homologous" to the organism in its untransformed state. Thus, lactose-dependent *E. Coli*. chromosomal DNA comprises a lactose or "lac" operon which mediates lactose digestion by, *inter alia*, elaborating the enzyme β -galactosidase. In the particular instance illustrated, the lac control elements are obtained from a bacteriophage, λ plac 5, which is infective for the *E. Coli*. The phage's lac operon, in turn, was derived by transduction from the same bacterial species, hence the "homology". Homologous regulons suitable for use in the disclosed process may alternatively derive from plasmidic DNA native to the organism.

The simplicity and efficiency of the lac promoter-operator system commend its use in the systems we describe, as does its ability to be induced by IPTG (isopropylthio- β -D galactoside). Of course, other operons or portions thereof could be employed as well, e.g., lambda promoter-operator, arabinose operon (phi 80 *dara*), the colicine EI, galactose, alkaline phosphatase or tryptophan operons. Promoter-operators derived from the latter (i.e., "tryp operon") would be expected to confer 100% repression pending induction (with indoleacrylic acid) and harvest.

5. Plasmid Construction Generally

The details of the process schematically illustrated in Figure 4 appear from the Experimental section, *infra*. At this point, however, it is useful to briefly discuss various of the techniques employed in constructing the recombinant plasmid of the preferred embodiment.

The cloning and expression of the synthetic somatostatin gene employed two plasmids. Each

plasmid has an *EcoRI* substrate site at a different region of the β -galactosidase structural gene (see Figures 4 and 5). The insertion of the synthetic somatostatin DNA fragment into the *EcoRI* sites of these plasmids brings the expression of the genetic information in that fragment under control of the lac operon controlling elements. Following the insertion of the somatostatin fragment into these plasmids, translation should result in a somatostatin polypeptide preceded either by 10 amino acid (pSOM1) of by virtually the whole β -galactosidase subunit structure (pSOM11—3).

The plasmid construction scheme initiates with plasmid pBR322, a well-characterized cloning vehicle. Introduction of the lac elements to this plasmid was accomplished by insertion of a *HaeIII* restriction endonuclease fragment (203 nucleotides) carrying the lac promoter, CAP binding site, operator, ribosome binding site, and the first 7 amino acid codons of the β -galactosidase structural gene. The *HaeIII* fragment was derived from λ plac5 DNA. The *EcoRI*-cleaved pBR322 plasmid, which had its termini repaired with T4 DNA polymerase and deoxyribonucleotide triphosphates, was blunt-end ligated to the *HaeIII* fragment to create *EcoRI* termini at the insertion points. Joining of these *HaeIII* and repaired *EcoRI* termini generate the *EcoRI* restriction site (see Fig. 4 and 5) at each terminus. Transformants of *E. Coli* RR1 with this DNA were selected for resistance to tetracycline (Tc) and ampicillin (Ap) on 5-bromo-4-chloro-incolylgalactoside (X-gal) medium. On this indicator medium, colonies constitutive for the synthesis of β -galactosidase, by virtue of the increased number of lac operators titrating repressor, are identified by their blue color. Two orientations of the *HaeIII* fragment are possible but these were distinguished by the asymmetric location of an *HhaI* restriction site in the fragment. Plasmid pBH10 was further modified to eliminate the *EcoRI* endonuclease site distal to the lac operator (pBH20).

The eight chemically synthesized oligodeoxyribonucleotides (Fig. 2) were labeled at the 5' termini with [³²P]-ATP by polynucleotide kinase and joined with T4 DNA ligase. Through hydrogen bonding between the overlapping fragments, the somatostatin gene self-assembles and eventually polymerizes into larger molecules because of the cohesive restriction site termini. The ligated products were treated with *EcoRI* and *BamHI* restriction endonucleases to generate the somatostatin gene as depicted in Figure 2.

The synthetic somatostatin gene fragment with *EcoRI* and *BamI* termini was ligated to the pBH20 plasmid, previously treated with the *EcoRI* and *BamHI* restriction endonucleases and alkaline phosphatase. The treatment with alkaline phosphatase provides a molecular selection for plasmids carrying the inserted fragment. Ampicillin-resistant transformants obtained with this ligated DNA were screened for tetracycline sensitivity and several were examined for the insertion of an *EcoRI*-*BamHI* fragment of the appropriate size.

Both strands of the *EcoRI*-*BamHI* fragments of plasmids from two clones were analyzed by nucleotide sequence analysis starting from the *BamHI* and *EcoRI* sites. The sequence analysis was extended into the lac controlling elements; the lac fragment sequence was intact, and in one case, pSOM1, the nucleotide sequence of both strands were independently determined each giving the sequence depicted in Figure 5A.

The *EcoRI*-*Pst* fragment of the pSOM1 plasmid, with the lac-controlling element, was removed and replaced with the *EcoRI*-*Pst* fragment of pBR322 to produce the plasmid pSOM11. The *EcoRI* fragment of λ plac 5, carrying the lac operon control region and most of the β -galactosidase structural gene, was inserted into the *EcoRI* site of pSOM11. Two orientations of the *EcoRI* lac fragment of λ plac 5 were expected. One of these orientations would maintain the proper reading frame into the somatostatin gene, the other would not. Analysis of independently isolated clones for somatostatin activity then identified clones containing the proper oriented gene, of which the clone designated pSOM11—3 was one.

6. The Microorganism

Various unicellular microorganisms have been proposed as candidates for transformation, such as bacteria, fungi and algae. That is, those unicellular organisms which are capable of being grown in cultures or fermentation. Bacteria are for the most part the most convenient organisms to work with. Bacteria which are susceptible to transformation include members of the Enterobacteriaceae, such as strains of *Escherichia coli* and *Salmonella*; Bacillaceae, such as *Bacillus subtilis*; *Pneumococcus*; *Streptococcus*, and *Haemophilus influenzae*.

The particular organism chosen for the somatostatin work next discussed was *E. Coli*. strain RR1, genotype: Pro⁻Leu⁻Thi⁻Thi⁻R_B⁻M_B⁻r_h⁻c A⁺ Str^r Lac^y *E. Coli*. RR1 is derived from *E. Coli*. HB 101 (H. W. Boyer, et al, J. Mol. Biol. (1969) 47, 459—472) by mating with *E. Coli* K12 strain KL16 as the Hfr donor. See J. H. Miller, Experiments in Molecular Genetics (Cold Spring Harbor, New York, 1972). Cultures of both *E. Coli* RR1 and *E. Coli* RR1 (pBR322) have been deposited with the American Type Culture Collection without restriction as to access, respectively ATCC Nos. 31343 and 31344. The somatostatin-producing organism has likewise been deposited [ATCC No. 31447].

In the case of human insulin, A and B chain genes were cloned in *E. Coli* K—12 strain 294 (end A, thi⁻, hsr⁻, hsm_K⁺), ATCC No. 31446, and that organism employed in expression of the A chain (*E. Coli* K—12 strain 294 [pAI], ATCC No. 31448). The B chain of human insulin was first expressed in a derivative of HB101, i.e., *E. Coli* K—12 strain D1210 a lac⁺ (i⁹⁰z⁺y⁺), and that B gene-containing

organism has likewise been deposited (ATCC No. 31449). Alternatively, the B gene may be inserted in and expressed from the organism first mentioned, i.e., strain 294."

EXPERIMENTAL

I SOMATOSTATIN

5 1. Construction of Somatostatin Gene Fragments 5

Eight oligodeoxyribonucleotides respectively labeled A through H in Figure 2 were first constructed, principally by the modified triester method of K. Itakura et al, *J. Am. Chem. Soc.* 97, 7327. (1975). However, in the case of fragments C, E and H resort was had to an improved technique in which

- 10 The improved technique is schematically depicted in Figure 3, wherein B is thymine, N-benzoylated adenine, N-benzoylated cytosine or N-isobutyrylated guanine. In brief, and with reference to Figure 3, 10
with an excess of I (2 mmole), the coupling reaction with II (1 mmole) went almost to completion in 60 min with the aid of a powerful coupling reagent, 2,4,6-trisopropylbenzenesulfonyl tetrazolide (TPSTe, 4 mmole; 2). After removal of the 5'-protecting group with 2% benzene sulfonic acid solution, the 5'-
15 hydroxyl dimer V could be separated from an excess of 3'-phosphodiester monomer IV by simple 15
solvent extraction with aqueous NaHCO₃ solution in CHCl₃. The fully protected trimer block was prepared successively from the 5'-hydroxyl dimer V, I (2 mmole), and TPSTe (4 mmole) and isolated by chromatography on silica gel, as in B. T. Hunt et al, *Chem. and Ind.* 1967, 1868 (1967). The yields of trimers made according to the improved technique appear from Table II.

TABLE II

Yields of Fully Protected Trimers

| Sequence | Yield | Sequence | Yield |
|----------|-------|----------|-------|
| TTT | 81% | ATG | 69% |
| TTT | 75% | GCC | 61% |
| GGA | 41% | CCA | 72% |
| AGA | 49% | CAA | 72% |
| ATC | 71% | TTA | 71% |
| CCT | 61% | CAT | 52% |
| ACA | 63% | CCC | 73% |
| ACC | 65% | AAC | 59% |
| CGT | 51% | GAT | 60% |

The eight oligodeoxyribonucleotides, after removal of all protecting groups, were purified by high-pressure liquid chromatography on Permaphase AAX (R. A. Henry et al *J. Chrom. Sci.* 11, 358 (1973)). The purity of each oligomer was checked by homochromatography on thin-layer DEAE-cellulose and also by gel electrophoresis in 20% acrylamide slab after labeling of the oligomers with [λ -³²P]-ATP in the presence of polynucleotide kinase. One major labeled product was obtained from each DNA fragment.

2. Ligation and Acrylamide Gel Analysis of Somatostatin DNA

- The 5' OH termini of the chemically synthesized fragments A through H were separately phosphorylated with T4 polynucleotide kinase. [λ -³²P]- γ -ATP was used in phosphorylation so that reaction products could be monitored autoradiographically, although it will be appreciated that unlabelled ATP would serve as well were autoradiographically with. Just prior to the kinase reaction, 25 μ Ci of [λ -³²P]ATP (approx. 1500 Ci/mMol) (Maxam and Gilbert, *Proc. Nat. Acad. Sci. U.S.A.* 74, 1507 (1977)) was evaporated to dryness in 0.5 ml Eppendorf tubes. Five micrograms of fragment were incubated with 2 units of T4 DNA kinase (hydroxylapatite fraction, 2500 units/ml; 27), in 70 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol in a total volume of 150 μ l for 20 min at 37°C. To insure maximum phosphorylation of the fragments for ligation purposes, 10 μ l of a mixture consisting of 70 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol, 0.5 mM ATP and two units of DNA kinase were added and incubation continued for an additional 20 min at 7°C. The fragments (250 ng/ μ l) were stored at -20°C without further treatment. Kinased fragments A, B, E, and F (1.25 μ g each) were ligated in a total volume of 50 μ l in 20 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM ATP and 2 units of T4 DNA ligase (hydroxylapatite fraction, 400 units/ml; 27), for 16 hr at 4°C. Fragments C, D, G and H were ligated under similar conditions. Samples of 2 μ l were removed for analysis by electrophoresis on a 10% polyacrylamide gel followed by autoradiography (H. L. Heyneker et al, *Nature* 263, 748 (1976)) in which unreacted DNA fragments are represented by fast migrating material and wherein the monomeric form of the ligated fragments migrate with bromophenol blue dye (BPB). Some dimerization also occurs by reason of the cohesive ends of the ligated fragments A, B, E and F, and of the ligated fragments C, D, G and H. These dimers represent the slowest migrating material, and may be cleaved by restriction endonuclease EcoRI and BamHI, respectively.

The two half molecules (ligated A + B + E + F and ligated C + D + G + H) were joined by an

additional ligation step carried out in a final volume of 150 μ l at 4°C for 16 hr. One microliter was removed for analysis. The reaction mixture was heated for 15 min at 65°C to inactivate the T4 DNA ligase. The heat treatment does not effect the migration pattern of the DNA mixture. Enough restriction endonuclease *Bam*HI was added to the reaction mixture to cleave the multimeric forms of the somatostatin DNA in 30 min at 37°C. After the addition of NaCl to 100 mM, the DNA was digested with *Eco*RI endonuclease. The restriction endonuclease digestions were terminated by phenol-chloroform extraction of the DNA. The somatostatin DNA fragment was purified from unreacted and partially ligated DNA fragments by preparative electrophoresis on a 10% polyacrylamide gel. The band containing the somatostatin DNA fragment was excised from the gel and the DNA was eluted by slicing the gel into small pieces and extracting the DNA with elution buffer (0.5 M ammonium acetate, 10 mM MgCl₂, 0.1 mM EDTA, 0.1% SDS) overnight at 65°C. The DNA was precipitated with 2 volumes of ethanol, centrifuged, redissolved in 200 μ l 10 mM Tris-HCl pH 7.6 and dialyzed against the same buffer resulting in a somatostatin DNA concentration of 4 μ g/ml.

3. Construction of Recombinant Plasmids

Figure 4 schematically depicts the manner in which recombinant plasmids comprising the somatostatin gene were constructed, and may be referred to in connection with the following more particularized discussion.

A. The Parental Plasmid pBR322

The plasmid chosen for experimental somatostatin cloning was pBR322, a small (molecular wt. approx. 2.6 megadaltons) plasmid carrying resistance genes to the antibiotics ampicillin (Ap) and tetracycline (Tc). As indicated in Figure 4, the ampicillin resistance gene includes a cleavage site for the restriction endonuclease *Pst*I, the tetracycline resistance gene includes a similar site for restriction endonuclease *Bam*HI, and an *Eco*RI site is situated between the Ap^r and Tc^r genes. The plasmid pBR322 is derived from pBR313, a 5.8 megadalton Ap^rTc^rCol^{imm} plasmid (R. L. Rodriguez et al, ICN—UCLA Symposia on Molecular and Cellular Biology 5, 471—77 (1976), R. L. Rodriguez et al, Construction and Characterization of Cloning Vehicles, in *Molecular Mechanisms in the Control of Gene Expression*, pp. 471—77, Academic Press, Inc. (1976). Plasmid pBR322 is characterized and the manner of its derivation fully described in F. Bolivar et al, "Construction and Characterization of New Cloning Vehicles II. A Multipurpose Cloning System", *Gene* (November 1977).

30 B. Construction of Plasmid pBH10

Five micrograms of plasmid pBR322 DNA was digested with 10 units of the restriction endonuclease *Eco*RI in 100 mM Tris-HCl pH 7.6, 100 mM NaCl, 6 mM MgCl₂ at 37°C for 30 min. The reaction was terminated by phenol-chloroform extraction; the DNA was then precipitated with two and a half volumes of ethanol and resuspended in 50 μ l of T4 DNA polymerase buffer (67 mM Tris-HCl pH 8.8, 6.7 mM MgCl₂, 16.6 mM (NH₄)₂SO₄, 167 μ g/ml bovine serum albumin, 50 μ M of each of the dNTP's; A. Panet et al, *Biochem.* 12, 5045 (1973). The reaction was started by the addition of 2 units of T4 DNA polymerase. After incubation for 30 min at 37° the reaction was terminated by a phenol-chloroform extraction of the DNA followed by precipitation with ethanol. Three micrograms of λ plac5 DNA (Shapiro et al *Nature* 224, 768 (1969)) was digested for 1 hr at 37°C with the restriction enzyme *Hae*III (3 units) in 6 mM Tris-HCl pH 7.6, 6 mM MgCl₂, 6 mM β -mercaptoethanol in a final volume of 20 μ l. The reaction was stopped by heating for 10 min at 65°C. The pBR322 treated DNA was mixed with the *Hae*III digested λ plac5 DNA and blunt-end ligated in a final volume of 30 μ l with 1.2 units of T4 DNA ligase (hydroxylapatite fraction; A. Panet et al, *supra*) in 20 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM ATP for 12 hrs at 12°C. The ligated DNA mixture was dialyzed against 10 mM Tris-HCl pH 7.6, and used for transformation of *E. coli* strain RR1. Transformants were selected for tetracycline and ampicillin resistance on minimal medium, plates containing 40 μ g/ml of 5-bromo-4-chloro-colygalactoside (X-gal) medium (J. H. Miller, Experiments in Molecular Genetics (Cold Spring Harbor, New York, 1972)). Colonies constitutive for the synthesis of β -galactosidase were identified by their blue color. After screening 45 independently isolated blue colonies, three of them were found to contain plasmid DNA carrying two *Eco*RI sites separated by approximately 200 base pairs. The position of an asymmetrically located *Hha*I fragment in the 203 b.p. *Hae*III lac control fragment (W. Gilbert et al, in Protein-Ligand Interactions, H. Sand and G. Blauer, Eds. (De Gruyter, Berlin, (1975) pp. 193—210) allows for the determination of the orientation of the *Hae*III fragment, now an *Eco*RI fragment, in these plasmids. Plasmid pBH10 was shown to carry the fragment in the desired orientation, i.e., lac transcription going into the Tc^r gene of the plasmid.

C. Construction of Plasmid pBH20

Plasmid pBH10 was next modified to eliminate the *Eco*RI site distal to the lac operator. This was accomplished by preferential *Eco*RI endonuclease cleavage at the distal site involving partial protection by RNA polymerase of the other *Eco*RI site localized between the Tc^r and lac promoters, which are only about 40 base pairs apart. After binding RNA polymerase, the DNA (5 μ g) was digested with *Eco*RI (1 unit) in a final volume of 10 μ l for 10 min at 37°C. The reaction was stopped by heating at 65°C for 10

min. The *EcoRI* cohesive termini were digested with SI nuclease in 25 mM Na-acetate pH 4.5, 300 mM NaCl, 1 mM ZnCl₂ at 25°C for 5 min. The reaction mixture was stopped by the addition of EDTA (10 mM final) and Tris-HCl pH 8 (5 mM final). The DNA was phenol-chloroform extracted, ethanol precipitated and resuspended in 100 μ l of T4 DNA ligation buffer. T4 DNA ligase (1 μ l) was added and the mixture incubated at 12°C for 12 hr. The ligated DNA was transformed in *E. coli* strain RR1, and Ap^r Tc^r transformants were selected on X-gal-antibiotic medium. Restriction enzyme analysis of DNA screened from 10 isolated blue colonies revealed that these clones carried plasmid DNA with one *EcoRI* site. Seven of these colonies had retained the *EcoRI* site located between the lac and Tc^r promoters. The nucleotide sequence from the *EcoRI* site into the lac-control region of one of these plasmids, pBH20, was confirmed. This plasmid was next used to clone the somatostatin gene.

D. Construction of Plasmid pSOM 1

Twenty micrograms of the plasmid pBH20 was digested to completion with restriction endonucleases *EcoRI* and *BamHI* in a final volume of 50 μ l. Bacterial alkaline phosphatase was added (0.1 unit of Worthington BAPF) and incubation was continued for 10 min at 65°C. The reactions were terminated by phenol-chloroform extraction and the DNA was precipitated with 2 volumes of ethanol, centrifuged and dissolved in 50 μ l 10 mM Tris-HCl pH 7.6, 1 mM EDTA. The alkaline phosphatase treatment effectively prevents self-ligation of the *EcoRI*, *BamHI* treated pBH20 DNA, but circular recombinant plasmids containing somatostatin DNA can still be formed upon ligation. Since *E. coli* RR1 is transformed with very low efficiency by linear plasmid DNA, the majority of the transformants will contain recombinant plasmids. Fifty microliters of somatostatin DNA (4 μ g/ml) were ligated with 25 μ l of the *BamHI*, *EcoRI*, alkaline phosphatase-treated pBH20 DNA in a total volume of 50 μ l containing 20 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 10 mM dithiothreitol, 0.5 mM ATP, and 4 units of T4 DNA ligase at 22°C. After 10, 20 and 30 min, additional somatostatin DNA (40 ng) was added to the reaction mixture (the gradual addition of somatostatin DNA may favor ligation to the plasmid over self-ligation). Ligation was continued for 1 hr followed by dialysis of the mixture against 10 mM Tris-HCl pH 7.6. In a control experiment, *BamHI*, *EcoRI*, alkaline phosphatase-treated pBH20 DNA was ligated in the absence of somatostatin DNA under similar conditions. Both preparations were used without further treatment to transform *E. coli* RR1. The transformation experiments were carried out in a P3 physical containment facility. (National Institutes of Health, U.S.A., Recombinant DNA Research Guidelines, 1976). Transformants were selected on minimal medium plates containing 20 μ g/ml Ap and 40 μ g/ml X-gal. Ten transformants, which were all sensitive to Tc, were isolated. For reference these were designated pSOM1, pSOM2, etc. . . . pSOM10. In the control experiment no transformants were obtained. Four out of the ten transformants contained plasmids with both an *EcoRI* site and *BamHI* site. The size of the small *EcoRI*, *BamHI* fragment of these recombinant plasmids was in all four instances similar to the size of the *in vitro* prepared somatostatin DNA. Base sequence analysis according to Maxam and Gilbert *Proc. Nat. Acad. Sci. U.S.A.* 74, 560 (1977), revealed that the plasmid pSOM1 had the desired somatostatin DNA fragment inserted.

The DNA sequence analysis of the clone carrying plasmid pSOM1 predicts that it should produce a peptide comprising somatostatin. However no somatostatin radioimmune activity has been detected in extracts of cell pellets or culture supernatants, nor is the presence of somatostatin detected when the growing culture is added directly to 70% formic acid and cyanogen bromide. *E. coli* RR1 extracts have been observed to degrade exogenous somatostatin very rapidly. The absence of somatostatin activity in clones carrying plasmid pSOM 1 could well result from intracellular degradation by endogenous proteolytic enzymes. Plasmid pSOM 1 was accordingly employed to construct a plasmid coding for a precursor protein comprising somatostatin and sufficiently large as to be expected to resist proteolytic degradation.

E. The Construction of Plasmids pSOM 11 and pSOM 11—3

A plasmid was constructed in which the somatostatin gene could be located at the C-terminus of the β -galactosidase gene, keeping the translation in phase. The presence of an *EcoRI* site near the C-terminus of this gene and the available amino acid sequence of this protein (B. Polisky et al, *Proc. Nat. Acad. Sci. U.S.A.* 73, 3900 (1976), A. V. Fowler et al, *Id.* at 74, 1507 (1976), A. I. Bukhari et al, *Nature New Biology* 243, 238 (1973) and K. E. Langley, *J. Biol. Chem.* 250, 2587 (1975)) permitted insertion of the *EcoRI* *BamHI* somatostatin gene into the *EcoRI* site while maintaining the proper reading frame. For the construction of such a plasmid, pSOM1 DNA (50 μ g) was digested with the restriction enzymes *EcoRI* and *PstI* in a final volume of 100 μ l. A preparative 5% polyacrylamide gel was used to separate the large *PstI*-*EcoRI* fragment that carries the somatostatin gene from the small fragment carrying the lac control elements. The large band was excised from the gel and the DNA eluted by slicing the gel into small pieces and extracting the DNA at 65°C overnight. In a similar way plasmid pBR322 DNA (50 μ g) was digested with *PstI* and *EcoRI* restriction endonucleases and the two resulting DNA fragments purified by preparative electrophoresis on a 5% polyacrylamide gel. The small *PstI*-*EcoRI* fragment from pBR322 (1 μ g) was ligated with the large *PstI*-*EcoRI* DNA fragment (5 μ g) from pSOM1 in a final volume of 50 μ l with 1 unit of T4 DNA ligase at 12°C for 12 hrs. The ligated mixture was used to transform *E. coli* RR1, and transformants were selected for ampicillin resistance on X-gal medium. As

expected, almost all the Ap^r transformants (95%) gave white colonies (no lac operator) on X-gal indicator plates. The resulting plasmid, pSOM11, was used in the construction of plasmid pSOM11—3. A mixture of 5 ug of pSOM11 DNA and 5 ug of λ plac⁵ DNA was digested with *Eco*RI (10 units for 30 min at 37°C). The restriction endonuclease digestion was terminated by phenol-chloroform extraction. The DNA was then ethanol-precipitated and resuspended in T4 DNA ligase buffer (50 μ l). T4 DNA ligase (1 unit) was added to the mixture and incubated at 12°C for 12 hrs. The ligated mixture was dialyzed against 10 mM Tris-HCl pH 7.6 and used to transform *E. Coli* strain RR1. Transformants were selected for Ap^r on X-gal plates containing ampicillin and screened for constitutive β -galactosidase production. Approximately 2% of the colonies were blue (pSOM11—1, 11—2 etc.). Restriction enzyme analysis of plasmid DNA obtained from these clones revealed that all the plasmids carried a new *Eco*RI fragment of approximately 4.4 megadaltons, which carries the lac operon control sites and most of the β -galactosidase gene. Because two orientations of the *Eco*RI fragment are possible, the asymmetric location of a *Hind*III restriction site was used to determine which of these colonies were carrying this *Eco*RI fragment with lac transcription proceeding into the somatostatin gene. *Hind*III-*Bam*HI double digestions indicated that only the clones carrying plasmids pSOM11—3, pSOM11—5, pSOM11—6 and pSOM11—7 contained the *Eco*RI fragment in this orientation.

4. Radioimmune Assay for Somatostatin Activity

The standard radioimmune assays (RIA) for somatostatin (A. Arimura et al, *Proc. Soc. Exp. Biol. Med.* 148, 784 (1975)) were modified by decreasing the assay volume and using phosphate buffer. Tyr¹¹ somatostatin was iodinated using a chloramine T procedure. (Id.) To assay for somatostatin, the sample, usually in 70% formic acid containing 5 mg/ml of cyanogen bromide was dried in a conical polypropylene tube (0.7 ml, Sarstedt) over moist KOH under vacuum. Twenty microliters of PBSA buffer (75 mM NaCl; 75 mM sodium phosphate, pH 7.2; 1 mg/ml bovine serum albumin; and 0.2 mg/ml sodium azide) was added, followed by 40 μ l of a [¹²⁵I] somatostatin "cocktail" and 20 μ l of a 1,000-fold dilution in PBSA of rabbit antisomatostatin immune serum S39 (Vale et al, *Metabolism* 25, 1491 (1976). The [¹²⁵I] somatostatin cocktail contained per ml of PBSA buffer: 250 ug normal rabbit gamma globulin (Antibodies, Inc.), 1500 units protease inhibitor ("Trasyolol", Calbiochem) and about 100,000 counts of [¹²⁵I] Tyr¹¹-somatostatin. After at least 16 hour at room temperature, 0.333 ml of goat anti-rabbit gamma globulin (Antibodies, Inc., P=0.03) in PBSA buffer was added to the sample tubes. The mixture was incubated 2 hr at 37°C, cooled to 5°C, then centrifuged at 10,000 \times g for 5 min. The supernatant was removed and the pellet counted in a gamma counter. With the amount of antiserum used, 20% of the counts was precipitated with no unlabeled competing somatostatin. The background with infinite somatostatin (200 ng) was usually 3%. One-half maximum competition was obtained with 10 pg of somatostatin. Initial experiments with extracts of *E. Coli* strain RR1 (the recipient strain) indicated that less than 10 pg of somatostatin could easily be detected in the presence of 16 μ g or more of cyanogen bromide-treated bacterial protein. More than 2 μ g of protein from formic acid-treated bacterial extracts interfered somewhat by increasing the background, but cyanogen bromide cleavage greatly reduced this interference. Reconstruction experiments showed that somatostatin is stable in cyanogen bromide-treated extracts.

40 A. Competition by Bacterial Extracts

Strains *E. Coli* RR1 (pSOM11—5) and *E. Coli* RR1 (pSOM11—4) were grown at 37°C to 5 \times 10⁸ cells/ml in Luria broth. Then IPTG was added to 1 mM and growth continued for 2 hr. One-milliliter aliquots were centrifuged for a few seconds in an Eppendorf centrifuge and the pellets were suspended in 500 μ l of 70% formic acid containing 5 mg/ml cyanogen bromide. After approximately 24 hr at room temperature, aliquots were diluted tenfold in water and the volumes indicated in Figure 6A were assayed in triplicate for somatostatin. In Figure 6A "B/B₀" is the ratio of [¹²⁵I] somatostatin bound in the presence of sample to that bound in the absence of competing somatostatin. Each point is the average of triplicate tubes. The protein content of the undiluted samples was determined to be 2.2 mg/ml for *E. Coli* RR1 (pSOM11—5) and 1.5 mg/ml for *E. Coli* RR1 (pSOM—4).

50 B. The Initial Screening of pSOM11 Clones for Somatostatin

Cyanogen bromide-treated extracts of 11 clones (pSOM11—2, pSOM11—3, etc.) were made as described above for the case of Figure 6A. Thirty microliters of each extract was taken in triplicate for radioimmune assay, whose results appear from Figure 6B. The range of assay points is indicated. The values for picograms somatostatin were read from a standard curve obtained as part of the same experiment.

* * *

The radioimmune assay results described thus far may be summarized as follows. In contrast to the results of experiments with pSOM1, four clones (pSOM11—3, 11—5, 11—6, and 11—7) were found to have easily detectable somatostatin radioimmune activity Figures 6A and 6B. Restriction fragment analysis revealed that pSOM11—3, pSOM11—5, pSOM11—6 and pSOM11—7 had the

desired orientation of the lac operon, whereas pSOM11—2 and 11—4 had the opposite orientation. Thus there is a perfect correlation between the correct orientation of the lac operon and the production of somatostatin radioimmune activity.

C. Effects of IPTG Induction and CNBr Cleavage on Positive and Negative Clones

The design of the somatostatin plasmid predicts that the synthesis of somatostatin would be under the control of the lac operon. The lac repressor gene is not included in the plasmid and the recipient strain (*E. coli* RR1) contains the wild type chromosomal lac repressor gene which produces only 10 to 20 repressor molecules per cell (15). The plasmid copy number (and therefore the number of lac operators) is approximately 20—30 per cell, so complete repression is impossible. As shown in Table III, *infra* the specific activity of somatostatin in *E. coli* RR1 (pSOM11—3) was increased by IPTG, an inducer of the lac operon. As expected, the level of induction was low, varying from 2.4 to 7 fold. In experiment 7 (Table III) α activity, a measure of the first 92 amino acids of β -galactosidase, also was induced by a factor of two. In several experiments no detectable somatostatin radioimmune activity can be detected prior to cyanogen bromide cleavage of the total cellular protein. Since the antiserum used in the radioimmune assay, S 39, requires a free N-terminal alanine, no activity was expected prior to cyanogen bromide cleavage.

TABLE III

Somatostatin Radioimmune Specific Activity

Abbreviations: Luria Broth, LB; isopropylthiogalactoside, IPTG; cyanogen bromide, CNBr; somatostatin, SS. Protein was measured by the method of Bradford, *Anal. Biochem.* 72, 248 (1976).

| Experiment Number | Strain | Medium | IPTG 1 mM | CNBr 5 mg/ml | pg SS/ug protein |
|-------------------|--------|----------------|-----------|--------------|------------------|
| 1 | 11—2 | LB | + | + | 0.1 |
| | 11—3 | LB | + | + | 12 |
| | 11—4 | LB | + | + | 0.4 |
| | 11—5 | LB | + | + | 15 |
| 2 | 11—3 | LB | + | + | 12 |
| | 11—3 | LB | + | — | 0.1 |
| 3 | 11—3 | LB | + | + | 61 |
| | 11—3 | LB | — | + | 8 |
| | 11—3 | LB | + | — | 0.1 |
| 4 | 11—3 | LB | + | + | 71 |
| | 11—3 | VB + glycerol* | + | + | 62 |
| | 11—3 | LB + glycerol | + | + | 250 |
| 5 | 11—3 | LB | + | + | 320 |
| 6 | 11—3 | LB | + | + | 0.1 |
| | 11—2 | LB | + | + | 24 |
| | 11—3 | LB | — | + | 10 |

* Vogel-Bonner minimal medium plus glycerol.

D. Gel Filtration of Cyanogen Bromide-Treated Extracts

Formic acid and cyanogen-treated extracts of the positive clones (pSOM 11—3, 11—5, 11—6, and 11—7) were pooled (Total volume 250 μ l), dried, and resuspended in 0.1 ml of 50% acetic acid. [3 H] leucine was added and the sample was applied to an 0.7 \times 47 cm column of Sephadex G—50 in 50% acetic acid. Fifty-microliter aliquots of the column fractions were assayed for somatostatin. Pooled negative clone extracts (11—2, 11—4, and 11—11) were treated identically. The results appear from Figure 5C. On the same column known somatostatin (Beckman Corp.) elutes as indicated (SS). In this system, somatostatin is well-separated from excluded large peptides and fully included small molecules. Only extracts of clones positive for somatostatin exhibited radioimmune activity in the column fractions and this activity elutes in the same position as chemically synthesized somatostatin.

SUMMARY OF ACTIVITY INFORMATION

The data establishing the synthesis of a polypeptide containing the somatostatin amino acid sequence are summarized as follows: (1) Somatostatin radioimmune activity is present in *E. coli* cells having the plasmid pSOM11—3, which contains a somatostatin gene of proven correct sequence and has the correct orientation of the lac *EcoRI* DNA fragment. Cells with the related plasmid pSOM11—2, which has the same somatostatin gene but an opposite orientation of the lac *EcoRI* fragment, produce no detectable somatostatin activity; (2) As predicted by the design scheme, no detectable somatostatin radioimmune activity is observed until after cyanogen bromide treatment of the cell extract; (3) The somatostatin activity is under control of the lac operon as evidenced by induction by IPTG, an inducer of the lac operon; (4) The somatostatin activity co-chromatographs with known somatostatin on Sephadex

G—50; (5) The DNA sequence of the cloned somatostatin gene is correct. If translation is out of phase, a peptide will be made which is different from somatostatin at every position. Radioimmune activity is detected indicating that a peptide closely related to somatostatin is made, and translation must be in phase. Since translation occurs in phase, the genetic code dictates that a peptide with the exact sequence of somatostatin is made; (6) Finally, the above samples of *E. coli* RR1 (pSOM11—3) extract inhibit the release of growth hormone from rat pituitary cells, whereas samples of *E. coli* RR1 (pSOM11—2) prepared in parallel and with identical protein concentration have no effect on growth hormone release.

STABILITY, YIELD, AND PURIFICATION OF SOMATOSTATIN

- The strains carrying the *Eco*RI lac operon fragment (pSOM11—2, pSOM11—3, etc.) segregate with respect to the plasmid phenotype. For example, after about 15 generations, about one-half of the *E. coli* RR1 (pSOM11—3) culture was constitutive for β -galactosidase, i.e., carried the lac operator, and of these about half were ampicillin resistant. Strains positive (pSOM11—3) and negative (pSOM11—2) for somatostatin are unstable, and therefore, the growth disadvantage presumably comes from the overproduction of the large but incomplete and inactive galactosidase. The yield of somatostatin has varied from 0.001 to 0.03% of the total cellular protein (Table 1) probably as the result of the selection for cells in culture having plasmids with a deleted lac region. The highest yields of somatostatin have been from preparations where growth was started from a single ampicillin resistant, constitutive colony. Even in these cases, 30% of the cells at harvest had deletions of the lac region. Storage in the frozen state (lyophilization) and growth to harvest from a single such colony is accordingly indicated for the system described. Yields may be increased by, e.g., resort to bacterial strains which overproduce lac repressor such that expression of precursor protein is essentially totally repressed prior to induction and harvest. Alternatively, as previously discussed, a tryptophan or other operator-promoter system which ordinarily is totally repressed may be employed.
- In the crude extract resulting from cell disruption in, e.g., an Eaton Press, the β -galactosidase-somatostatin precursor protein is insoluble and is found in the first low speed centrifugation pellet. The activity can be solubilized in 70% formic acid, 6M guanidinium hydrochloride, or 2% sodium dodecyl sulfate. Most preferably, however, the crude extract from the Eaton Press is extracted with 8M urea and the residue cleaved with cyanogen bromide. In initial experiments somatostatin activity derived from *E. coli* strain RR1 (pSOM 11—3) has been enriched approximately 100-fold by alcohol extraction of the cleavage product and chromatography on Sephadex G—50 in 50% acetic acid. When the product is again chromatographed on Sephadex G—50 and then subjected to high pressure liquid chromatography, substantially pure somatostatin may be obtained.

II. HUMAN INSULIN

- The techniques previously described were next applied to the production of human insulin. Thus, the genes for insulin B chain (104 base pairs) and for insulin A chain (77 base pairs) were designed from the amino acid sequence of the human polypeptides, each with single-stranded cohesive termini for the *Eco*RI and *Bam*HI restriction endonucleases and each designed for insertion separately into pBR322 plasmids. The synthetic fragments, deca- to pentadeca-nucleotides, were synthesized by the block phosphotriester method using trinucleotides as building blocks and ultimately purified with high performance liquid chromatography (HPLC). The human insulin A and B chain synthetic genes were then cloned separately in plasmid pBR322. The cloned synthetic genes were fused to an *E. Coli* β -galactosidase gene as before to provide efficient transcription, translation, and a stable precursor protein. Insulin peptides were cleaved from β -galactosidase precursor, detected by radioimmunoassay, and purified. Insulin radioimmunoassay activity was then generated by mixing the *E. Coli* products.

1. Design and Synthesis of Human Insulin Genes

- The genes constructed for human insulin are depicted in Figure 9. The genes for human insulin, B chain and A chain, were designed from the amino acid sequences of the human polypeptides. The 5' ends of each gene have single stranded cohesive termini for the *Eco*RI and *Bam*HI restriction endonucleases, for the correct insertion of each gene into plasmid pBR322. A *Hind*III endonuclease recognition site was incorporated into the middle of the B chain gene for the amino acid sequence Glu-Ala to allow amplification and verification of each half of the gene separately before the construction of the whole B chain gene. The B chain and the A chain genes were designed to be built from 29 different oligodeoxyribonucleotides, varying from decamer to pentadecamers. Each arrow indicates the fragment synthesized by the improved phosphotriester method, H1 to H8 and B1 to B12 for the B chain gene and A1 to A11 for the A chain gene.

2. Chemical Synthesis of Oligodeoxyribonucleotides

- Materials and methods for synthesis of oligodeoxyribonucleotides were essentially those described in Itakura, K. et al (1975) *J. Biol. Chem.* 250, 4592 and Itakura, K. et al (1975) *J. Amer. Chem. Soc.* 97, 7327 except for these modifications:
- The fully protected mononucleotides, 5'-O-dimethoxytrityl-3'-*p*-chlorophenyl- β -cyanoethyl

phosphates, were synthesized from the nucleoside derivatives using the monofunctional phosphorylating agent *p*-chlorophenyl- β -cyanoethyl phosphorochloridate (1.5 molar equivalent) in acetonitrile in the presence of 1-methyl imidazole Van Boom, J. H. et al (1975) *Tetrahedron* 31, 2953. The products were isolated in large scale (100 to 300g) by preparative liquid chromatography (Prep 500 LC, Waters Associates).

b) By using the solvent extraction method [Hirose, T. et al (1978) *Tetrahedron Letters*, 2449] 32 bifunctional trimers were synthesized (see Table IV) in 5 to 10 mmole scale, and 13 trimers, 3 tetramers, and 4 dimers as the 3' terminus blocks, in 1 mmole scale. The homogeneity of the fully protected trimers was checked by thin layer chromatography on silica gel in two methanol/chloroform solvent systems: solvent a, 5% v/v and solvent b, 10% v/v (See Table IV). Starting from this library of compounds, 29 oligodeoxyribonucleotides of defined sequence were synthesized, 18 for the B chain and 11 for the A chain gene.

The basic units used to construct polynucleotides were two types of trimer block, i.e. the bifunctional trimer blocks of Table IV and corresponding 3'-terminus trimers protected by an anisoyl group at 3'-hydroxy. The bifunctional trimer was hydrolyzed to the corresponding 3'-phosphodiester component with a mixture of pyridine-triethylamine-water (3:1:1 v/v) and also to the corresponding 5'-hydroxyl component with 2% benzenesulfonic acid. The 3'-terminus block previously referred to was treated with 2% benzenesulfonic acid to give the corresponding 5'-hydroxyl. The coupling reaction of an excess of the 3'-phosphodiester trimer (1.5 molar equivalent) with the 5'-hydroxyl component, however obtained, (1 molar equivalent) in the presence of 2,4,6-triisopropylbenzenesulfonyl tetrazolide (TPSTe, 3 to 4 equivalents) went almost to completion diester block reactant the reaction mixture was passed through a short silica gel column set up on a sintered glass filter. The column was washed, first with CHCl_3 to elute some side products and the coupling reagent, and then with CHCl_3 :MeOH (95:5 v/v) in which almost all of the fully protected oligomer was eluted. Under these conditions, the charged 3'-phosphodiester block reactant remained in the column. Similarly, block couplings were repeated until the desired length was constructed.

High performance liquid chromatography (HPLC) was used extensively during oligonucleotide synthesis for a) analysis of each trimer and tetramer block, b) analysis of the intermediate fragments (hexamers, monomers, and decamers), c) analysis of the last coupling reaction, and d) purification of the final products. The HPLC was performed by using a Spectra-Physics 3500B liquid chromatograph. After removal of all protecting groups by conc. NH_4OH at 50°C (6 h) and 80% AcOH at room temperature (15 min), the compounds were analyzed on a Permaphase AAX (DuPont) [Van Boom, J. et al (1977) *J. Chromatography* 131, 169.] column (1 m x 2 mm), using a linear gradient of solvent B (0.05M KH_2PO_4 - 1.0M KCl, pH 4.5) in solvent A (0.01M KH_2PO_4 , pH 4.5). The gradient was formed starting with buffer A and applying 3% of buffer B per minute. The elution was performed at 60°C, with a flow rate of 2 ml per minute. The purification of the 29 final oligonucleotides also was performed on Permaphase AAX, under the same conditions reported above. The desired peak was pooled, desalted by dialysis, and lyophilized. After labeling the 5' termini with (γ - ^{32}P)ATP using T^4 polynucleotide kinase, the homogeneity of each oligonucleotide was checked by electrophoresis on a 20% polyacrylamide gel.

TABLE IV
SYNTHESIS OF TRIMER BUILDING BLOCKS

| No. | Compound* | Yield** (%) | a. | Rf b. | Purity*** (%) | In Figure 9, Present In: |
|-----|-----------|----------------|------|----------|------------------|-----------------------------|
| 1. | AAG | 47 | 0.15 | 0.40 | 93 | B5, B6 |
| 2. | AAT | 49 | 0.25 | 0.52 | 95 | H1, A1, A6 |
| 3. | AAC | 52 | 0.28 | 0.55 | 93 | H5, B6, A2, A8 |
| 4. | ACT | 43 | 0.27 | 0.53 | 91 | B4, B5, A6 |
| 5. | ACC | 56 | 0.33 | 0.60 | 96 | B7 |
| 6. | ACG | 39 | 0.18 | 0.45 | 90 | H5, B7 |
| 7. | AGG | 45 | 0.10 | 0.26 | 89 | H6, H7, B9 |
| 8. | AGT | 33 | 0.14 | 0.40 | 96 | B9, A2, A11 |
| 9. | AGC | 50 | 0.19 | 0.48 | 92 | H8, B1, A5, A10 |
| 10. | AGA | 48 | 0.24 | 0.50 | 91 | A9 |
| 11. | TTC | 44 | 0.26 | 0.52 | 95 | B4, B7, A3 |
| 12. | TTC | 49 | 0.11 | 0.31 | 94 | H3, H5, A2, A3, A5 |
| 13. | JCT | 58 | 0.24 | 0.49 | 96 | A4 |
| 14. | TCA | 45 | 0.28 | 0.53 | 92 | H1, H2, H4, A1 |
| 15. | TCG | 39 | 0.12 | 0.34 | 91 | A2 |
| 16. | TGG | 32 | 0.10 | 0.28 | 87 | H3, A1, A10 |
| 17. | TGC | 51 | 0.18 | 0.47 | 93 | H6, B2, A4, A7, A8 |
| 18. | TGA | 46 | 0.12 | 0.37 | 94 | H7 |
| 19. | TAC | 61 | 0.22 | 0.50 | 90 | B4, A11 |
| 20. | TAA | 55 | 0.17 | 0.44 | 95 | B5, A10 |
| 21. | CCT | 53 | 0.30 | 0.55 | 97 | H3, H4, B10 |
| 22. | CAC | 47 | 0.25 | 0.51 | 92 | A3 |
| 23. | CAA | 58 | 0.25 | 0.51 | 93 | H2, H6, H8, A7 |
| 24. | CTT | 41 | 0.28 | 0.54 | 92 | B6, B9, A4 |
| 25. | CGA | 40 | 0.27 | 0.52 | 93 | A7 |
| 26. | CGT | 75 | 0.25 | 0.50 | 89 | H2, H4, B3, B1 |
| 27. | GGT | 35 | 0.09 | 0.26 | 90 | B3 |
| 28. | GTT | 46 | 0.18 | 0.45 | 93 | B2 |
| 29. | GTA | 38 | 0.25 | 0.50 | 95 | B6, B8, A6 |
| 30. | GAA | 39 | 0.15 | 0.39 | 88 | H7, B3, B8, A5 |
| 31. | GAT | 52 | 0.22 | 0.49 | 89 | B10, A9 |
| 32. | GCA | 42 | 0.14 | 0.39 | 93 | A9 |

* Fully protected trideoxynucleotides; 5'-O-(2,4,6-trimethylbenzyl)-3'-O-(4-chlorophenyl)-β-D-ribofuranose-2-phosphate

** Yield was the overall yield calculated from the 5'-hydroxylmonomers.

*** Based on HPLC analysis.

3. Assembly and Cloning of B Chain Gene and the A Chain Gene

The gene for the B chain of insulin was designed to have an *EcoRI* restriction site on the left end, a *HindIII* site in the middle and *BamHI* site at the right end. This was done so that both halves, the left *EcoRI-HindIII* half (BH) and the right *HindIII-BamHI* half (BB), could be separately cloned in the convenient cloning vehicle pBR322 and after their sequences had been verified, joined to give the complete B gene (Figure 10). The BB half was assembled by ligation from 10 oligodeoxyribonucleotides, labeled B1 to B10 in Figure 9, made by phosphotriester chemical synthesis. B1 and B10 were not phosphorylated, thereby eliminating unwanted polymerization of these fragments through their cohesive ends (*HindIII* and *BamHI*). After purification by preparative acrylamide gel electrophoresis and elution of the largest DNA band, the BB fragment was inserted into plasmid pBR322 which had been cleaved with *HindIII* and *BamHI*. About 50% of the ampicillin resistant colonies derived from the DNA were sensitive to tetracycline, indicating that a nonplasmid *HindIII-BamHI* fragment had been inserted. The small *HindIII-BamHI* fragments from four of these colonies (pBB101 to pBB104) were sequenced and found to be correct as designed.

The BH fragment was prepared in a similar manner and inserted into pBR322 which had been cleaved with *EcoRI* and *HindIII* restriction endonucleases. Plasmids from three ampicillin resistant, tetracycline sensitive transformants (pBH1 to pBH3) were analyzed. The small *EcoRI-HindIII* fragments were found to have the expected nucleotide sequence.

The A chain gene was assembled in three parts. The left four, middle four, and right four oligonucleotides (see Figure 9) were ligated separately, then mixed and ligated (oligonucleotides A1 and A12 were unphosphorylated). The assembled A chain gene was phosphorylated, purified by gel electrophoresis, and cloned in pBR322 at the *EcoRI-BamHI* sites. The *EcoRI-BamHI* fragments from two ampicillin resistant, tetracycline sensitive clones (pA10, pA11) contained the desired A gene sequence.

4. Construction of Plasmids for Expression of A and B Insulin Genes

Figure 10 illustrates the construction of the lac-insulin B plasmid (pIB1). Plasmids pBH1 and

pBB101 were digested with *EcoRI* and *HindIII* endonucleases. The small BH fragment of pBH1 and the large fragment of pBB101 (containing the BB fragment and most of pBR322) were purified by gel electrophoresis, mixed, and ligated in the presence of *EcoRI*-cleaved λ plac5. The megadalton *EcoRI* fragment of λ plac5 contains the *lac* control region and the majority of the β -galactosidase structural gene. The configuration of the restriction sites ensures correct joining of BH to BB. The *lac EcoRI* fragment can insert in two orientations; thus, only half of the clones obtained after transformation should have the desired orientation. The orientation of ten ampicillin resistant, β -galactosidase constitutive clones were checked by restriction analysis. Five of these colonies contained the entire B gene sequence and the correct reading frame from the β -galactosidase gene into the B chain gene. One, PIB1, was chosen for subsequent experiments.

In a similar experiment, the 4.4 megadalton *lac* fragment from λ plac5 was introduced into the pA11 plasmid at the *EcoRI* site to give pIA1. pIA1 is identical to pIB1 except that the A gene fragment is substituted for the B gene fragment. DNA sequence analysis demonstrated that the correct A and B chain gene sequences were retained in pIA1 and pIB1 respectively.

15 5. Expression 15

The strains which contain the insulin genes correctly attached to β -galactosidase both produced large quantities of a protein the size of β -galactosidase. Approximately 20% of the total cellular protein was this β -galactosidase-insulin A or B chain hybrid. The hybrid proteins are insoluble and were found in the first low speed pellet where they constitute about 50% of the protein.

To detect the expression of the insulin A and B chains, we used a radioimmunoassay (RIA) based on the reconstitution of complete insulin from the separate chains. The insulin reconstitution procedure of Katsoyannis et al (1967) *Biochemistry* 6, 2642—2655, adapted to a 27-microliter assay volume, provides a very suitable assay. Easily detectable insulin activity is obtained after mixing and reconstituting S-Sulfonated derivatives of the insulin chains. The separated S-sulphonated chains of insulin do not react significantly, after reduction and oxidation, with the anti-insulin antibody used.

To use the reconstitution assay we partially purified the β -galactosidase-A or B chain hybrid protein, cleaved with cyanogen bromide, and formed S-sulfonated derivatives.

The evidence that we have obtained correct expression from chemically synthesized genes for human insulin can be summarized as follows: a) Radioimmune activity has been detected for both chains. b) The DNA sequences obtained after cloning and plasmid construction have been directly verified to be correct as designed. Since radioimmune activity is obtained, translation must be in phase. Therefore, the genetic code dictates that peptides with the sequences of human insulin are being produced. c) The *E. coli* products, after cyanogen bromide cleavage, behave as insulin chains in three different chromatographic systems which separate on different principles (gel filtration, ion exchange, and reversed phase HPLC). d) The *E. coli* produced A chain has been purified on a small scale by HPLC and has the correct amino acid composition.

CLAIMS:

1. A recombinant plasmid suited for transformation of a bacterial host and use therein as a cloning vehicle, wherein the plasmid comprises:

- a) A regulon homologous to the bacterial host in its untransformed state; and
- b) In reading phase with the regulon, a DNA insert coding for the amino acid sequence of a heterologous polypeptide, such that bacteria transformed by the plasmid are capable of expressing said amino acid sequence in recoverable form.

2. A bacterial plasmid according to claim 1.

3. A plasmid according to claim 2 whose regulon is essentially identical to a regulon ordinarily present in the bacterial host's chromosomal DNA.

4. A plasmid according to claim 2 wherein the regulon comprises the promoter-operator system of an operon selected from the group consisting of *E. coli* lac operon and *E. coli* tryptophan operon.

5. A plasmid according to either claim 3 or claim 4 wherein the heterologous polypeptide is a mammalian hormone or an intermediate therefor.

6. A plasmid according to claim 5 wherein the heterologous polypeptide is selected from the group consisting of human preproinsulin, human proinsulin, the A chain of human insulin, the B chain of human insulin, human or bovine growth hormone, leutinizing hormone, ACTH and pancreatic polypeptide.

7. The plasmid of claim 6 wherein the heterologous polypeptide is a mammalian hormone.

8. The plasmid of claim 5 wherein the heterologous polypeptide is somatostatin.

9. The plasmid of claim 8 wherein the regulon comprises the promoter-operator system of the *E. coli* lac operon.

10. A plasmid according to claim 1, substantially as described in Experimental I or Experimental II herein.

TAB FF

12

EUROPEAN PATENT APPLICATION

21 Application number: 79301054.7

51 Int. Cl.3: C 12 P 21/02, C 12 N 15/00

22 Date of filing: 04.06.79

30 Priority: 08.06.78 US 913533

43 Date of publication of application: 09.01.80
Bulletin 80/1

24 Designated Contracting States: AT BE CH DE FR GB IT
LU NL SE

71 Applicant: PRESIDENT AND FELLOWS OF HARVARD
COLLEGE, Massachusetts Avenue, Cambridge
Massachusetts 02138 (US)

72 Inventor: Gilbert, Walter, 107 Upland Road, Cambridge
Massachusetts 02140 (US)
Inventor: Broome, Stephanie A., 17 Chillon Street,
Cambridge Massachusetts 02140 (US)
Inventor: Villa-Komaroff, Lydia J., 15 Pond Circle,
Boston, Massachusetts 02130 (US)
Inventor: Efstratiadis, Argiris A., 321 Harvard Street,
Cambridge Massachusetts 02138 (US)

74 Representative: Harrison, David Christopher et al,
MEWBURN ELLIS & CO 70 & 72 Chancery Lane,
London WC2A 1AD (GB)

54 Method of making a selected protein.

57 A plasmid or phage gene for a periplasmic or extra-
cellular bacterial protein is cleaved, a doublestranded DNA
sequence coding for a selected protein or portion thereof
from a eukaryotic cell such as insulin is inserted in that
cleaved gene by recombinant DNA techniques and used to
transform a bacterium, and the excreted selected protein is
collected.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479

-1-

METHOD OF MAKING A SELECTED PROTEIN

The invention described herein was made in the course of work under a grant or award from the United States Department of Health, Education and Welfare.

5 TECHNICAL FIELD OF INVENTION

This invention relates to a process for producing specific proteins in bacteria and having them excreted from the bacterial cell and pertains more specifically to inserting the DNA representing the desired non-bacterial protein or part of a protein by recombinant techniques into the plasmid or phage gene for either a periplasmic or an extracellular protein, hereinafter called a "carrier protein", transforming a bacterial host with the recombined gene, and culturing the transformed host to excrete the protein. The protein thus produced can be collected by conventional procedures from the culture medium or from the periplasmic space depending upon the choice of carrier protein gene.

BACKGROUND ART

20 It is known to insert DNA representing a specific protein into the gene for an intracellular protein. Scheller *et al.*, Science, 196, 177-180 (1977). The problem however is that proteins made in this way are mixed with other intracellular proteins and are therefore subject to degradation by enzymes within the cell so that there is a problem in obtaining the desired protein product in purified form.

DISCLOSURE OF INVENTION

30 The foregoing problem is avoided in accordance with our invention by providing a method of making a selected protein or portion thereof by inserting DNA representing the selected protein or portion thereof into a bacterial gene, characterized by cleaving the bacterial gene for an extracellular or periplasmic carrier protein, inserting into the cleavage site by a recombinant step a non-bacterial

35

DNA fragment which codes for the selected protein or portion thereof, transforming a bacterial host with the recombined gene, and culturing the transformed bacteria to excrete the selected protein or portion thereof.

5 By way of example, by employing a gene for a carrier protein which has a leader sequence of hydrophobic amino acids at its amino terminus and which is normally excreted through the membrane of the cell within which it is made, with cleavage of the hydrophobic
10 leader sequence during excretion, a selected protein can be produced which can be recovered either from the periplasmic space or from the medium in which the bacterium is grown, depending upon the choice of carrier protein. In this way contamination from the other
15 proteins within the bacterium is avoided while achieving greater stability by avoiding the enzymes within the bacterial cell which degrade foreign proteins.

Among the bacterial genes for carrier proteins which can be employed in the present invention are the
20 genes for antibiotic resistance, such as the gene for penicillin resistance or penicillinase, the gene for chloramphenicol resistance, or the gene for tetracycline resistance, as well as the gene for alkaline phosphatase and the gene for bacterial ribonuclease.

25 Genes or DNA fragments which code for a variety of proteins or portions thereof can be inserted in the bacterial carrier protein gene by the process of the present invention. These proteins include a variety of non-bacterial proteins such as eukaryotic cell proteins and viral proteins. Of parti-
30 cular importance are eukaryotic cell proteins such as insulin, human growth hormone, interferon and other pharmacologically active proteins. These are synthesized by their respective genes as pre-proteins or precursor proteins having at their amino terminus a series of
35 hydrophobic amino acids. This hydrophobic leader sequence is not identical to that for the bacterial

proteins which are excreted through the bacterial membrane. Therefore, the fact that pre-insulin or other pre-proteins of higher cells contain a hydrophobic leader sequence is in itself no basis for expecting that such a pre-protein could be matured in the bacterial cell even if it could be synthesized within the cell. Moreover, the process of the present invention, in addition to providing for the synthesis within and excretion from bacterial cells of matured proteins of eukaryotic cells, which are of known utility, also makes possible the synthesis in and excretion from bacterial cells of other extracellular products of commercial interest. These include other fused proteins and fused proteins consisting of carrier proteins, as defined above, which carry specific determinants, for example, viral antigens such as coat proteins or other antigenic proteins of viruses. These latter fused proteins are useful in the manufacture of vaccines, being capable because of their antigenic character of inducing generation of an immune response specific to the viruses. Such vaccines will be unusually safe because they will not contain any live or inactivated virus material. Furthermore, it is possible by this process to construct vaccines for viruses which cannot be grown in culture.

25 BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1, 2 and 3 of the drawing show the complete base sequence for the E. coli penicillinase gene carried on the plasmid pBR322 along with the corresponding amino acid sequence of the protein for which it codes.

30 BEST MODE OF CARRYING OUT THE INVENTION

The following specific example is intended to illustrate more fully the nature of the present invention without acting as a limitation upon its scope.

Example

35 There was employed as the carrier protein E. coli [Escherichia coli] penicillinase, the gene for which is carried on the

small plasmid pBR322. A restriction enzyme map of this gene is shown in the drawing. This plasmid vector has been described by Bolivar et al., Gene, 2, 95-113 (1977). As the host bacteria we employed E. coli 1776*; see Curtiss et al. in Recombinant Molecules: Impact on Science and Society, Proceedings of the Tenth Miles International Symposium, eds. Beers & Bassett, 45-56 (1977). The host-vector combination is a certified EK2 system, certified by the United States National Institute of Health (NIH), July 7, 1977.

The plasmid carries a Pst [providencia stuartii endonuclease] restriction site of the penicillinase gene corresponding to the position of amino acids 181 and 182, as shown in the drawing. Double stranded cDNA was synthesized from RNA containing preproinsulin in RNA (PPI-mRNA) isolated from an X-ray induced, transplantable rat B-cell tumor (Chick et al., Proc. Natl. Acad. Sci. USA [P.N.A.S.], 74, 628-632 [1977]), Batches of 20g each of frozen tumor slices were ground with sterile sand with mortar and pestle and the cytoplasmic RNA purified from a post-nuclear supernatant by Mg^{2+} precipitation (Palmiter, Biochemistry, 13, 3603-3615 (1974)) followed by extraction with phenol and chloroform. This RNA was further purified by aligo-dT-cellulose chromatography (Aviv et al., P.N.A.S., 69, 1408-1412 (1972)) and used directly as template for double-stranded cDNA synthesis, as described (Efstratiadis et al., Cell, 7, 279-288 (1976)), except that a specific (dpT)₈ dpGpC primer (Collaborative Research) was utilized for reverse transcription. The concentrations of RNA and primer were 7 mg/ μ l and 1 mg/ μ l, respectively. All four α -³²P-dNTPs were at 1.25 mM (final specific activity 0.85 Ci/m mole). The reverse transcript was 2% of the input RNA, and 25% of it was finally recovered in the double-stranded DNA product.

*A deposit of the *Escherichia coli* X 1776 has been placed with and made available to the public at the American Type Culture Collection, Rockville, Maryland, U.S.A. and has been assigned ATCC No. 31244.

The double-stranded cDNA was inserted into the Pst site of plasmid pBR322 by the following procedure: pBR322 DNA (5.0 µg) was linearized with Pst and approximately 15 dG residues were added per 3' end by terminal transferase at 15°C in the presence of 1 mM Co²⁺ (Roychoudhury et al., Nucleic Acids Res., 3,101-116(1976)) and 100 µg/ml autoclaved gelatin. The same procedure was used to add dC residues to 2.0 µg of double-stranded cDNA. The reaction mixtures were extracted with phenol, and ethanol precipitated. The dC tailed double-stranded cDNA was electrophoresed in a 6% polyacrylamide gel under native conditions. Following autoradiography, molecules in the size range of 300 to 600 base pairs (.5 µg) were eluted from the gel (Efstratiadis et al. in Methods in Molecular Biology, 8,1-124 (1976)). The eluted double-stranded cDNA was concentrated by ethanol precipitation, redissolved in 10 mM Tris pH 8, mixed with 5 µg dG tailed pBR322 and dialized versus 0.1 M NaCl, 10 mM EDTA, 10 mM Tris pH 8. The mixture (4 ml) was then heated at 56° for 2 minutes, and annealing was performed at 42° for 2 hours. The hybrid DNA was used to transform E. coli 1776. The use of oligo dC-dG joins regenerates the Pst cuts so that the insert may be later excised.

Transformation of E. coli x1776 (an EK-2 host with pBR322, an EK-2 vector) was performed in a biological safety cabinet in a P3 physical containment facility in compliance with N.I.H. guidelines for recombinant DNA research published in the Federal Register, July 7, 1976. x1776 was transformed by a transfection procedure (Enea et al., J.Mol.Biol., 96, 495-509 (1975)) slightly modified as follows: x1776 was grown in L Broth [10 gms tryptone, 5 gm yeast extract, 5 gm NaCl (Difco)] supplemented with 10 µg/ml diaminopimelic acid and 40 µg/ml thymidine (Sigma) to A₅₅₀ of 0.5. A 200 ml portion of cells were sedimented at 1500 rpm and

resuspended by swirling in 1/10 volume of cold buffer containing 70 mM MnCl_2 , 40 mM NaAc pH 5.6, 30 mM CaCl_2 and kept on ice for 20 minutes. The cells were repelleted and resuspended in 1/30 of the original volume in the same buffer. The annealed DNA (2 ml) was added to the cells. Aliquots of this mixture (0.3 ml) were placed in sterile tubes and incubated on ice 60 minutes. The cells were then placed at 37° for 2 minutes. Broth was added to each tube (.7 ml) and the tubes incubated at 37° for 15 minutes. A 200 μl portion of the cells was spread on sterile nitrocellulose filters (Millipore) overlaying agar plates containing 15 $\mu\text{g/ml}$ tetracycline. (The filters were boiled to remove detergents before use.) The plates were incubated at 37° for 48 hours. Replicas of the filters were made. The nitrocellulose filters containing the transformants were removed from the agar and placed on a layer of sterile Whatman filter paper. A new sterile filter was placed on top of the filter containing the colonies and pressure was applied with a sterile velvet cloth and a duplicate block. A sterile needle was used to key the filters. The second filter was placed on a new agar plate and incubated at 37° for 48 hr. The colonies on the first filter were screened by the Grunstein-Hogness (P.N.A.S., 72, 3961-3965 (1975)) technique, using as probe an 80-nucleotide long fragment produced by Hae III digestion of high specific activity cDNA copied from the rat oligo-dT bound RNA. Positive colonies were rescreened by the HART method (Paterson *et al.*, P.N.A.S., 74, 4370-4374 (1977)) as follows: Plasmid DNA (about 3 μg) was digested with Pst, ethanol precipitated, and dissolved directly into 20 μl dionized formamide. After heating for one minute at 95° each sample was placed on ice. After the addition of 1.5 μg oligo (dT)-cellulose bound RNA, PIPES at pH 6.4 to 10 mM and NaCl to 0.4 M, the mixtures were incubated for 2 hr at 50°. They were

then diluted by the addition of 75 μ l H₂O and ethanol precipitated in the presence of 10 μ g wheat-germ tRNA, washed with 70% ethanol, dissolved in H₂O and added to a wheat-germ cell-free translation mixture (Roberts et al., P.N.A.S., 70, 2330-2334 (1973)). After three hours at 23°C, duplicate 2 μ l aliquots were removed for trichloroacetic acid precipitation; the remainder of the reaction mixture was treated with ribonuclease, diluted with immunoassay buffer, and analyzed for the syntheses of immunoreactive preproinsulin by means of a double antibody immunoprecipitation (Lomedico et al. Nucleic Acids Res., 3, 381-391 (1976)). The washed immunoprecipitates were dissolved in 1 ml of NCS (Amersham) and counted in 10 μ l of Omnifluor (New England Nuclear) by liquid scintillation.

One colony was identified by the HART Screening. The Pst excisable insert was sequenced by the method of Maxam and Gilbert (P.N.A.S., 74, 560-564 (1977)) to show that it corresponded to the sequence of rat preproinsulin I. This insert, labeled by nick translation with DNA polymerase I was used to screen 200 transformants with the Grunstein-Hogness assay. There were identified 48 clones containing rat preproinsulin cDNA inserts.

These 48 clones of transformed E. coli x1776 were screened using an in situ radioimmunoassay technique to determine whether the clones were producing insulin antigens and whether they were producing fused polypeptide chains, one end of which being insulin antigen and the other end penicillinase (the bacterial carrier protein) antigen. Presence of the fused polypeptide chains would indicate that the clones contained genes which were the products of the fusion of the bacterial gene for penicillinase with the eukaryotic cell gene for insulin. Such fused polypeptide chains were in fact found, using the technique to be described below. The technique takes

5 advantage of the fact that the fused proteins being
searched for contain two antigenic ends, each of which
will bind to its respective specific antibody. A specific
antibody was laid down on a plastic disk, the antigenic
protein from lysed bacterial cells placed in contact
with this disk, then the disk was rinsed and exposed
to radioactive antibodies. A protein molecule will
bind to the antibody fixed to the plastic with one
antigenic determinants and will bind in turn a radio-
10 active antibody with a second determinant. If anti-
penicillinase is on the disk and anti-insulin is
labeled, after the "sandwich" is washed, the only points
of radioactivity remaining will mark the presence of
fused proteins. In more detail, the method was as
15 follows:

Each 8.25 cm diameter disk of clear polyvinyl
(PV) 8 mm thick (Dora May Co., New York) was flattened
between sheets of smooth paper. In a glass petri dish,
each disk was then placed upon the surface of a liquid
20 containing 10 ml of 0.2 MNaHCO₃, at pH 9.2, containing
60 µg/ml IgG. After 2 minutes or longer at room
temperature, the disk was removed and washed twice with
10 ml of cold wash buffer (WB), which consisted of
phosphate-buffered saline, 0.5% normal guinea pig serum,
25 0.1% bovine serum albumin and 0.3 mg/ml streptomycin
sulfate. Each disk was used immediately after washing.

Antigens were released from bacterial cells
by transferring colonies onto 1.5% agarose containing 0.5
mg lysozyme/ml, 30 mM Tris pH 8, and 10 mM EDTA. The
30 IgG-coated surface of a PV disk was placed face down on
the agarose and bacterial colonies and left for 60
minutes at 4°. Each disk was then removed and washed 3
times with 10 ml of cold WB. This step completed the
immunoabsorption of antigen onto the solid-phase antibody
35 layer.

Reaction of the ¹²⁵I-labeled antibodies with

the antigen now adhering to the disks was done by setting 1.5 ml WB containing 5×10^6 cpm (γ emission) ^{125}I -IgG onto the center of an 8.25 cm diameter flat disk of ordinary nylon mesh which had been placed in the bottom of a petri dish. The mesh served as a spacer. A disk treated as in the earlier steps then was placed facedown on the mesh and solution and incubated overnight at 4° . Each disk was then washed twice with 10 ml cold WB and twice with water, and allowed to dry at room temperature. At this point, fused proteins had bound to both the ordinary and radioactively labeled layers of IgG. These proteins were then detected with conventional autoradiography technique using Kodak No. Screen Film or Kodak X-OMAT R film and a DuPont Cronex Lighting plus intensifying screen as described for example by Laskey *et al.*, FEBS Lett., 82, 314-316 (1977). Both anti-insulin and anti-penicillinase IgG fractions were required for the procedure above. The anti-insulin antiserum was a commercially available product obtained from guinea pigs. The rabbit anti-penicillinase anti-serum was produced by injection (1 mg pure) penicillinase (in complete Freund's adjuvant (Difco)) into New Zealand white rabbits. (Booster injections were administered in incomplete Freund's adjuvant (Difco)) 2 and 3 weeks after the initial injection, and the rabbits were bled 1 week later.

The IgG fractions were prepared from each immune serum by ammonium sulfate precipitation followed by DEAE-cellulose (Whatman, DE-52) chromatography in 0.025 M potassium phosphate, pH 7.3, 1% glycerol. Fractions containing the bulk of the flow-through material were pooled, and protein was precipitated by adding ammonium sulfate to 40% saturation. The resulting pellet was resuspended in 1/3 the original serum volume of 0.025 M potassium phosphate, pH 7.3, 0.1 M NaCl, 1% glycerol, and dialyzed against the same buffer. After

dialysis, any residual precipitate was removed by centrifugation. IgG fractions were stored in aliquots at -70° .

Each IgG fraction was radioiodinated by the usual method of Hunter *et al.*, Biochem. J., 91, 43 - 46 (1964). The 25 μ l reaction mixture contained 0.5 M potassium phosphate, pH 7.5, 2 mCi carrier-free Na^{125}I , 150 μ g IgG and 2 μ g chloramine T. After 3 minutes at room temperature, 8 μ g of sodium metabisulfite in 25 μ l PBS was added, followed by 200 μ l PBS containing 2% normal guinea pig serum. The ^{125}I -labeled IgG was purified by chromatography on a Sephadex G-50 column equilibrated with PBS containing 2% normal guinea pig serum. The ^{125}I -IgG elution fraction was diluted to 5 ml with PBS containing 10% normal guinea pig serum, filtered through a sterile Millipore VC filter (0.1 μ m pore size), divided into aliquots and stored at -70° . The specific activities were 1.5×10^7 cpm/ μ g.

This screening detected one clone of x1776 that synthesized and secreted a fused protein showing both penicillinase and insulin antigenic determinant. This protein, recovered from the periplasm, mimics insulin in radioimmunoassays. DNA sequencing shows that this protein is a fusion between penicillinase and proinsulin, the two proteins being connected by 6 glycines between amino acid 182 of penicillinase (Alamine) and amino acid 4, glutamine, of proinsulin. Thus a higher cell hormone has been synthesized in bacteria in an antigenically active form.

It will be appreciated that the DNA sequence for the desired eukaryotic cell protein can be inserted into a Hind II cut corresponding to the position between amino acids 101 and 102 of the protein for which this pBR322 plasmid codes, or into the Taq cut at the position corresponding to amino acid 45. In all cases, if the eukaryotic cell DNA is arranged in phase, by the random addition of tails or by other procedures, it will

be expressed as a fused part of the carrier protein; and the protein excreted from the cell. Furthermore, the sequence of the penicillinase gene, as it exists in this plasmid, or in others, can be modified either by mutation, or by direct recombinant DNA techniques such as the insertion of DNA fragments at specific points within the gene, in such way as to insert new restriction cuts that are convenient for splicing. For example, the R1 cut on the plasmid pBR322 can be removed by mutation, and an R1 sequence inserted by ligation into the penicillinase gene. Although this might inactivate the gene, it would not interfere with the use of this region of DNA to synthesize a carrier protein.

The segment of the penicillinase gene DNA between the code for amino acid 23 at the end of the hydrophobic leader and the code for amino acid 45 at the Taq cut for example, can be removed by nibbling back the DNA by a mixture of appropriate enzymes. One such mixture is the lambda exonuclease which will chew back the DNA strand from the 5' end, together with the enzyme S1, which will remove the single stranded overhang. Another such mixture is T₄ DNA polymerase which will chew back the 3' end of one DNA strand together with S1, which again will remove the single stranded overhang. By controlled digestion the plasmid DNA molecule can be appropriately shortened to the fragment extending from the R1 cut to the point coding for amino acid 23 or to other points on the hydrophobic leader sequence, and such a fragment can be fused to a similarly generated fragment containing the insulin sequence, chewed back enzymatically to a convenient initial point, presumably again, the point where the mature insulin molecule begins. These two fragments can be fused together, for example, by butt end ligation by the T₄ DNA ligase, and that fusion inserted into the plasmid. That fusion produces a degenerate species of the carrier protein,

for which the carrier gene codes for only the E. coli hydrophobic leader sequence and the eukaryotic cell gene provides the rest of the structural information. Although such construction can in principle be done
5 exactly, in practice they will probably be done on a random basis, involving the splicing of a variety of gene fragments whose end points are in interesting regions, and examining the medium surrounding clones of bacteria transformed by the fused fragments to detect
10 antigenic activity by an RIA such as the one described above, as evidence of protein synthesis.

The procedure of the present invention is not restricted to the use of the E. coli penicillinase gene, but is applicable to the gene for any excreted
15 protein carried on a multicopy plasmid or on a phage. It is not restricted to insulin, but can be used to find the expression of the fused protein of any DNA fragment of a virus or eukaryotic cell that carries a coding region that codes, when translated in phase, for
20 antigenic determinants in the viral or eukaryotic cell protein. Thus if fragments of animal virus DNA are inserted into the Pst or Hind II site of the penicillinase gene, some recipient bacterium will synthesize a fused protein which will be recognizable by using
25 the RIA technique, employing antibodies specific to the viral antigen. This fused protein in turn can be purified and used to stimulate an antibody response in an animal or person, either for the production of antibodies directed at specific sites on the virus
30 protein, or as vaccination against the viral antigen. The fused protein will provide helper determinants in such a vaccination, to aid the immune response, although, presumably, aggregated states of the fused protein would have to be used in a vaccine. The specific carrier
35 proteins that would be used might be either the bacterial

-13-

proteins themselves or still further fusions between the bacterial proteins and other convenient sequences to provide useful helper determinants in carrier protein.

1. A method of making a selected protein or portion thereof by inserting DNA representing the selected protein or portion thereof into a bacterial gene, characterized by cleaving the bacterial gene for an extracellular or periplasmic carrier protein, inserting into the cleavage site by a recombinant step a non-bacterial DNA fragment which codes for the selected protein or portion thereof, transforming a bacterial host with the recombined gene, and culturing the transformed bacteria to excrete the selected protein or portion thereof.
2. The method according to claim 1, characterized in that the selected protein is a eukaryotic cell protein which contains a hydrophobic leader which is normally cleaved during excretion from the eukaryotic cell.
3. The method according to claim 1 or 2, characterized in that the carrier protein is E. coli penicillinase.
4. The method according to any of claims 1 to 3, characterized in that the selected protein is insulin.
5. The method according to claim 1, characterized in that the carrier protein is E. coli penicillinase and the bacterial gene is cleaved at the Pst restriction site.

R₁
 CAATTCTTGAAGACGAAGGGCCTCGTGATAGCGCTATTTTATAGGTTAAATGTCATGATAATAATGGTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTG
 CTTAAGAACTTCTGCTTTCGGGAGCACTATGCGGATAAAAATATCCAATTACAGTACTATTATTACCAAGAATCTGAGTCCACCGTGAAAAGCCCCCTTTACAC

CGGGAAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACATAACCCCTGATAAATGCTTCAATAATTTGAAAAAGGAAG
 GCGCCTTGGGGATAACAATAAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTAACTTTTTCCTTC

MetSerIleGlnHisPheArgValAlaLeuIleProPhePheAlaAlaPheCysLeuProValPheAlaHisProGluThrLeuValLysValLysAsp
 AGTATGAGTATCAACATTTCCGTTGCGCCTTATTCCTTTTGGGCAATTTGGCTCCCTGTTTTTGCTCACCACAGAAACGCTGGTGAAGTAAAGAT
 TCATACTCATAGTTGTAAGGCCACAGCGGGAAATAAGGGAAAAACCGCTAAACGGAGGACAAAAACGAGTGGGTCTTTGGACCACTTTCATTTTCTA

AlaGluAspGlnLeuGlyAlaArgValGlyTyrIleGluLeuAspLeuAsnSerGlyLysIleLeuGluSerPheArgProGluGluArgPheProMetMet
 GGTGAAGATCAGTTGGGTGGACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAGATCTCTTGAGAGTTTCGCCCCGGAAGAACGTTTCCCATGTG
 CGACTTCTAGTCAACCCACGTGCTCACCCTAATGTAGTTGACCTAGAGTTGTGCGCCATCTTAGGAACCTCTCAAAAGCGGGGCTTCTTGCAAAAGTTTACTAC

SerThrPheLysValLeuLeuCysGlyAlaValLeuSerArgValAspAlaGlyGlnGluGlnLeuGlyArgArgIleHisTyrSerGlnAsnAspLeuVal
 AGCACATTTTAAAGTCTGCTATGTGGCGCGGTATATCCCGTGTGACGCCCGGGCAAGAGCAACTCGGTGCGCGCATACACTATTTCTGAGAAIGACTTGGTT
 TCGTGAAAATTTCAAGACCGATACACCGCGCCATAATAGGGCACAACTGCGGCCCGTCTCGTTGAGCCAGCGGCGGTATGTGTAAGAGTCTTACTGAACCAA

FIG.1

0006694

110 120 130
 GluThrSerProValThrGluLysHisLeuThrAspGlyMetThrValArgGluLeuCysSerAlaIleThrMetSerAspAsnThrAlaIleAsnLeu
 GAGTACTCACCAAGTCACAGAAAGCATCTTACGGATGGCATGACAGTAAAGAAATTATGAGTGTGCTGCCATAACCATGAGTGAATAACACTGCGGCCAACTTA

 520 540 560 580 600
 CTCATGAGTGGTCAGTGTCTTTTCGTAGAAATGCCCTACCGTACTGTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACGCCGCTTGAAT

 140 150 160
 LeuLeuThrThrIleGlyGlyProLysGluLeuThrAlaPheLeuHisAsnMetGlyAspHisValThrArgLeuAspArgTrpGluProGluLeuAsnGlu
 CTTCTGACAAAGATCGAGGACCGAGGAGCTAACCGCTTTTGGACAACATGGGGGATCATGTAACTGCTTGGTATCGTTGGGAACCGGGAGCTGAATGAA

 620 640 660 680 700
 GAAGACTGTTGCTAGCTCCTGGCTTCCTCGATTGGCGAAACAGTGTGTACCCCTAGTACATTGAGGGGAACTAGCAACCCCTTGGCCCTCGACTTACTT

 170 180 190 200
 AlaIleProAsnAspGluArgAspThrThrMetProAlaIleMetAlaThrThrLeuArgLysLeuLeuThrGlyGluLeuLeuThrLeuAlaSerArgGln
 GCCATACCAACGACGAGCGTGAACACCGATGCCCTGCAGCAATGGCAACAGGTGGGCAACATTAACCTGGGGAACCTACTTACTCTAGCTTCCCGGCAA

 720 740 760 780 800
 CGGTATGGTTGCTGCTCGCACTGTGGTGCTACGGAGCTCGTTACCGTTGTTGCAACGCGTTTGATAATTGACCGCTTGATGAATGAGATCGAAGGGCGGTT

 210 220 230
 GlnLeuIleAspTrpMetGluAlaIleValAlaGlyProLeuLeuArgSerAlaLeuProAlaGlyTrpPheIleAlaAspLysSerGlyAlaGlyGln
 CAATTAAATAGACTGGATGGAGGCGGATAAGTGCAGGACCACTTCTGGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCCGGTGAG

 820 840 860 880 900
 GTTAATTATCTGACCTACCTCGGCCTATTTCAACGTCCTGGTGAAGACGCCGAGCCGGGAGCGCCGACCAATAACGACTATTAGACCTCGGCCACTC

 240 250 260 270
 ArgGlySerArgGlyIleIleAlaIleLeuGlyProAspGlyLysProSerArgIleValValIleLeuThrThrGlySerGlnAlaThrMetAspGluArg
 CGTGGGTCTCGCGGTATCATTTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGA

 920 940 960 980 1000 1020
 GCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTTGGGAGGGCATAGCATCAATAGATGTGTGTCGCCCTCAGTCCGTTGATACCTACTTGCT

FIG. 2

0006694

280 286
 AsnArgGlnIleAlaGluIleGlyAlaSerLeuIleLysHisTrp
 AATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAGCATTGGTAACGTGCAGACCAAGTTTACTCATATATATACTTTAGA
 1040 1060 1080 1100
 TTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTGTAACCATTTGACAGTCTGGTTCAAATGAGTATATATGAAATCT

3

FIG. 3

0006694

TAB GG

Ann. Rev. Microbiol. 1976. 30:507-33

Copyright © 1976 by Annual Reviews Inc. All rights reserved

GENETIC MANIPULATION OF MICROORGANISMS: POTENTIAL BENEFITS AND BIOHAZARDS

◆1691

Roy Curtiss III

Department of Microbiology, Institute of Dental Research, Cancer Research and
Training Center, University of Alabama, Birmingham, Alabama 35294

CONTENTS

| | |
|---|-----|
| INTRODUCTION | 507 |
| CURRENT STATUS OF TECHNOLOGY | 508 |
| POTENTIAL BENEFITS | 509 |
| <i>Introduction</i> | 509 |
| <i>Acquisition of Basic Knowledge</i> | 509 |
| <i>Applications of Technology</i> | 510 |
| POTENTIAL BIOHAZARDS | 511 |
| <i>Introduction</i> | 511 |
| <i>Factors Affecting the Probability of Escape of Chimeric Microorganisms</i> | 512 |
| <i>Factors Affecting the Probability of Survival and Perpetuation of Recombinant DNA</i> | 514 |
| <i>Factors Affecting Probable Consequences of Escape and Survival of Recombinant DNA Molecules</i> | 518 |
| <i>Estimating the Probability of Manifestation of Adverse Consequences by Recombinant DNA Molecules</i> | 519 |
| INTENTIONAL RELEASE OF GENETICALLY ALTERED MICROORGANISMS | 527 |
| <i>For Beneficial Purposes</i> | 527 |
| <i>For NonBeneficial Purposes</i> | 528 |
| COPING WITH POTENTIAL BIOHAZARDS | 528 |
| DATA NEEDED TO EVALUATE POTENTIAL BIOHAZARDS OF CHIMERIC MICROORGANISMS | 530 |
| CONCLUDING REMARKS | 530 |

INTRODUCTION

It has recently become possible to enzymatically couple viral, procaryotic, or eucaryotic DNA to bacteriophage or plasmid-cloning vector DNA and to introduce these recombinant DNA molecules into bacteria to clone and ultimately to allow

507

study of the expression of the foreign genetic information (15, 37). The potential applications of these techniques in obtaining basic information about life processes and in attempting to contend with some of the major problems facing our society stagger the mind (48, 74). On the other hand, the use of these newly acquired techniques for the genetic manipulation of microorganisms and their ultimate adaptation for the genetic manipulation of eucaryotic organisms is also associated with potential risks. In this review, I consider only briefly some of the potential beneficial applications of these techniques; instead, I dwell more comprehensively on the potential biohazards that are of more imminent concern and the means to contend with them. Considerations are restricted principally to the cloning of DNA in microorganisms (i.e. procaryotic organisms for which cloning techniques are already developed and other free-living, generally unicellular, lower eucaryotic organisms, such as algae, fungi, and protozoans, for which cloning techniques are not yet developed). Such genetically altered microorganisms that possess genetic information from species that do not normally exchange genetic material with the host microbe are referred to as chimeric microorganisms.

CURRENT STATUS OF TECHNOLOGY

Cohen (15) has recently reviewed the historical developments leading to the ability to clone foreign genes in the bacterium *Escherichia coli* K 12. In order to prepare recombinant DNA molecules in vitro, DNA is isolated from an organism of choice, either with or without suitable enrichment for specific gene sequences, and then one of two methods is used to construct recombinant DNA molecules. In one procedure, the DNA is fragmented either enzymatically or by shearing, then treated with an exonuclease that digests a single DNA strand in the 5' to 3' direction; then single-stranded homo-deoxypolymeric sequences are added to the 3' single-stranded ends of the molecule by using a suitable deoxyribonucleotide triphosphate (dXTP) species and terminal deoxynucleotidyl transferase (42, 51). The cloning vector DNA is similarly cleaved and treated with exonuclease, then complementary homodeoxypolymeric sequences are added by use of terminal transferase. The two DNA fractions are permitted to anneal and then may be treated with exonuclease III, DNA polymerase I, and polynucleotide-joining enzyme (DNA ligase) to form covalently closed circular recombinant DNA molecules (42, 51). In another procedure, the DNA is fragmented by using a restriction endonuclease which generates short complementary cohesive ends at both termini of the DNA duplex (35, 54, 69). The cloning vector DNA is treated with the same enzyme, the two preparations are mixed to permit annealing, and ligase is added to yield covalently closed circular recombinant DNA molecules (18, 79). Recombinant DNA molecules formed by either method are then used to transform a suitable bacterium (17, 52) to yield either transformants inheriting the chimeric plasmid or phage-containing cloned DNA segments. A variety of plasmid cloning vectors (18, 39, 79, 82) and bacteriophage λ variants (29, 59, 66, 81) have been developed that possess features that facilitate their use for cloning foreign DNA. Most research to date has utilized *E. coli* K 12 strains for the cloning of DNA from a staphylococcus plasmid (12), *Xenopus laevis*

(58), sea urchins (45), *Drosophila melanogaster* (32, 80, 81, 83), mouse mitochondria (13), yeast (J. Carbon, personal communication), and undoubtedly a greater diversity of other organisms by the time this article goes to press. The techniques can be readily adapted to use phage- or plasmid-cloning vectors that are able to replicate in procaryotic hosts other than *E. coli*. Sufficient information is also at hand to enable the use of various plant and animal DNA-containing viruses as vectors for cloning DNA in eucaryotic host cells infectable by the virus in question. Although so far unrealized, it is possible that plasmid DNA obtained from eucaryotic cell organelles will be usable as cloning vectors for eucaryotic cell research.

POTENTIAL BENEFITS

Introduction

The construction of chimeric microorganisms that possess and express specific genetic information from species that do not ordinarily exchange genetic information with the bacterial host presents certain problems which need to be solved. Although there should be no difficulty in observing expression of genes from procaryotic organisms of bacterial plasmid or phage DNA, it is unlikely that eucaryotic genetic information would be expressed often in procaryotic hosts. This problem makes it difficult to use random DNA segments from a eucaryotic organism to form recombinant DNA molecules (i.e. a "shotgun" experiment) and then to select a bacterial transformant possessing the desired eucaryotic genetic information. At present, such desired transformant clones can be obtained by starting with a DNA that has been either enriched for a specific DNA sequence or copied from a purified RNA (i.e. cDNA) and then recognized by using specific molecular probes, such as a purified RNA species (11). Alternatively, the starting material can be unfractionated DNA, and the desired clone can be recovered by using a statistical enrichment procedure in conjunction with the purified RNA probe (16). Thus, in order to facilitate cloning of any eucaryotic gene, it is necessary first to understand the basic differences between procaryotic and eucaryotic transcription and translation mechanisms and then to either couple eucaryotic genetic information with the appropriate start and stop signals of procaryotic origin or endow the procaryotic host with the ability to permit faithful transcription and translation of eucaryotic information. Endeavors are already in progress to couple specific procaryotic operator-promoter sequences onto eucaryotic DNA sequences, and it can be anticipated that the intensive research in this area will eventually solve these problems. It is also evident that the use of chimeric microorganisms to produce an abundance of certain desired gene products will mean that the host strains must be manipulated to liberate the product into the growth medium without cell lysis.

Acquisition of Basic Knowledge

The use and refinement of the techniques for constructing chimeric microorganisms would provide a wealth of knowledge that cannot easily be obtained by other means (48, 74). For example, the ability to clone specific eucaryotic DNA sequences would

permit characterization of eucaryotic chromosome structure, including the basis for maintenance of reiterated sequences, the distinctions between the DNA in euchromatic and heterochromatic chromosome regions (if any), and the eventual construction of complete chromosome maps. These techniques can also be used to deduce the mechanism for generating the potential for antibody diversity as well as other aspects of genetic regulatory phenomena, even in the absence of the ability to express such eucaryotic DNA. When the problem of expression of eucaryotic DNA in procaryotic organisms is solved, it would be possible to study gene regulation and do complementation analyses and fine-structure mapping of genes. These techniques would be useful in characterizing viral gene functions relevant to understanding infection, latency, and oncogenesis. Procaryotic genetic information from organisms that have unusual growth properties and that lack well-defined systems of genetic analysis can also be studied in a genetic background, which would permit all the tools of modern *E. coli* genetics to be used to elucidate the basis for these important metabolic activities. These new types of information on the structure, expression, and regulation of genetic information from a diversity of organisms will facilitate our understanding of both normal and abnormal functioning of life processes. The application of such new knowledge will have a profound effect on life and living and will undoubtedly lead to applications of the technology to the solution of problems that are completely unforeseen at present.

Applications of Technology

Although one cannot predict all applications of the ability to genetically engineer microorganisms, some applications are readily apparent, and others, although more speculative, can be predicted. Certainly, chimeric microorganisms could be used to produce products useful in medicine, agriculture, commerce, and industry. In the area of public health, I envision the construction of strains for the development of safe, effective vaccines to cope with viral, bacterial, mycotic, and parasitic infections, as well as the potential use of microbes to directly contend with obnoxious pathogens, insect pests, etc. The continued production of a sufficient quantity of high-quality foodstuffs to feed the population of the world is one of society's major goals, but despite considerable progress in this area, we do not appear to be any closer to attaining this goal. Therefore chimeric microorganisms will undoubtedly be used to produce large quantities of high-quality animal proteins inexpensively, for use as diet supplements, or, alternatively, to genetically manipulate the microorganism so as to eliminate problems associated with endotoxins and high concentrations of nucleic acids, thereby rendering them suitable sources of nutritive, nontoxic single-cell protein. The cloning of specific DNA sequences in microorganisms could also be used to genetically manipulate plant and animal species of agricultural importance to increase nutritive value and/or reduce reliance on expensive fertilizers or foodstuffs. Furthermore, these technologies would most likely be used to contend with specific types of environmental pollution and in some instances to degrade accumulated wastes through the production of usable, inexpensive energy sources.

In my view, the discoveries leading to the ability to couple genes from organisms that do not normally exchange genetic information and to introduce them into microorganisms where the foreign genes can be cloned and ultimately expressed

GENETIC MANIPULATION OF MICROORGANISMS 511

represents one of the most profound series of scientific discoveries in the modern era of science. Although I am extremely optimistic about the benefits that will accrue from this research, it should also be evident that the utilization of these skills to synthesize organisms that probably never existed previously is not without hazard. Certainly, the growth of large volumes of chimeric microorganisms, with the potential for their inadvertent release and even more their design for intentional release into the biosphere for a specific purpose, may have unpredictable consequences for the quality of life of a diversity of species inhabiting the biosphere.

POTENTIAL BIOHAZARDS

Introduction

The potential biohazards associated with recombinant DNA molecule research led to a series of discussions in 1973 (73) and 1974 that culminated in the publication of a letter by a NAS-NRC committee chaired by Paul Berg (8). This letter briefly assessed some of the potential risks and requested that scientists defer certain types of experiments and use considerable caution and good judgment in the performance of other types of experiments. This letter initiated extensive discussion among members of the scientific community that has gone on unabated since that time and has led to the convening of an international meeting at the Asilomar Conference Center in Pacific Grove, California, in February 1975 for the express purpose of considering all aspects of recombinant DNA molecule research and formulating provisional guidelines for the performance of this research (9). These topics have also been discussed and debated at a series of other scientific meetings held throughout the world since that time. These discussions also led to the formation of a National Institutes of Health Recombinant DNA Molecule Program Advisory Committee to establish more formal guidelines for US investigators. Similar committees are at work in other nations to devise comparable guidelines (2) so that this research can be conducted worldwide in a manner that would ensure maximum benefits while minimizing, if not precluding, manifestation of biohazards.

During these discussions and debates, some members of the scientific community have contended that the construction of recombinant DNA molecules and their introduction into microorganisms that might be inadvertently released from the laboratory environment would be without great potential biohazard. One such argument, which has considerable merit from a genetic standpoint, is that a bacterial cell endowed with additional traits that are without benefit to the cell would not compete well with cells that are not burdened with the maintenance and expression of this additional genetic information. The limitation of this argument is that there is no accurate way to predict whether a new trait will or will not confer a selective advantage to the cell in any of the various ecological niches in which the cell might find itself.

Another oft-heard argument is that such composites of genetic information have undoubtedly been tried during the course of evolution, with the consequence that such combinations have not been successful and therefore have not survived. Although I would agree that such combinations of traits might have occurred at some

time during the course of evolution, I reject the concept that these combinations of traits, when introduced into the biosphere as we know it now, would not pose a problem. Certainly, the earth has also evolved and conditions that now prevail are quite different from those present at the time when such combinations of traits might have been unsuccessfully tried.

Because there is a conspicuous lack of information about whether some of the surmised potential biohazards are real or not, some scientists have also taken the position that one should not worry about these potential biohazards until data are available to substantiate such speculations. Although I agree that it is uncharacteristic for scientists to fear the unknown or to postpone utilization of an available procedure to help deduce the secrets of life, it must be noted that chimeric microorganisms and the recombinant DNA which they contain have the potential for perpetuation and thus for irreversible fixation in living organisms. Since scientists are citizens of the world, it would therefore be totally irresponsible for any member of the scientific community to exercise any less than the required or advisable set of precautions for the performance of any given experiment in regard to the potential biohazards associated with it. I therefore consider it much more prudent to assume the worst and to investigate the existence and severity of postulated biohazards before relaxing the various safeguards for performing experiments on construction and use of microbes with recombinant DNA molecules.

In evaluating the potential biohazards associated with the construction of a chimeric microorganism, one must consider the probability of its escape from the research laboratory environment, the probability of survival and perpetuation of the recombinant DNA in nature, and lastly the probability that, if recombinant DNA molecules survive and are perpetuated in nature, adverse consequences could be manifested. The factors contributing to each of these probabilities and the possible variables need to be considered before attempting to assess what these probabilities may be.

Factors Affecting the Probability of Escape of Chimeric Microorganisms

PERSONNEL The first line of defense against the inadvertent release of a chimeric microorganism or of a viral vector carrying foreign DNA has to be the personnel who work in or around a laboratory in which such research is being conducted. All those actively engaged in the research must be both well trained in the use of required procedures, instruments and facilities, and meticulous practitioners of these skills. Furthermore, other laboratory workers such as dishwashers, media preparers, etc, and even occasional or intermittent laboratory visitors such as secretaries, custodians, equipment repair technicians, plumbers, electricians, etc, must all understand the nature of the experiments being conducted, and especially the nature of the associated biohazards, real or potential, so that any entry by these individuals into the research laboratory and any of their actions will not lead to the escape by chimeric microorganisms or viral vectors. The procedures and precautions all these individuals employ should be commensurate with the risks and should follow what is ordinarily associated with research on pathogenic microorganisms (14, 19, 26, 38, 47, 49, 60, 61, 70, 78).

GENETIC MANIPULATION OF MICROORGANISMS 513

Unfortunately, this is easier said than done. For example, although I am familiar with most of the procedures for working with virulent pathogens and although our laboratory sterilizes or decontaminates everything prior to disposal or dishwashing, many years of working on the genetics of *E. coli* K12 have led to the adoption of procedures that are totally unsuitable for working with chimeric microorganisms. For example, dishwashers do not automatically discard all bottle or tube caps lacking liners or screw cap tubes with chipped lips; screw cap centrifuge bottles are never inverted with fluid therein to check for leaks; gaskets on centrifuge rotors are not routinely replaced so as to maintain a good seal; and, although methods for correct sterilization of bacteriological loops and needles are known, they are seldom practiced. Because of these factors and because many individuals who are now interested in undertaking experiments on the formation of recombinant DNA molecules have never had any training in microbiological skills, there is some doubt about whether the experimentalist will actually adopt procedures that would effectively eliminate any inadvertent escape by chimeric microorganisms. Even when appropriate procedures are employed, laboratory workers are often known to acquire antibody titers against the virus or bacterial species with which they are working—a good indication that laboratory workers will be exposed to chimeric microorganisms, potentiating the possibility of the workers becoming the vectors for carrying microorganisms out of the laboratory environment.

Another personnel-related problem concerns the probability of contaminating cultures used to construct or containing chimeric microorganisms. In either case, recombinant DNA molecules, if inadvertently permitted to escape from the laboratory environment, may be introduced into microbial hosts that possess a higher probability for survival and potential for subsequent transmission of recombinant DNA than the intended host. In this regard, it is also evident that the likelihood of cultures used for recombinant DNA molecule research becoming contaminated with robust permissive strains would be increased if the same laboratory group is actively engaged in research with the microbial species and/or vectors being used for cloning.

FACILITIES Considerable knowledge is available about the use of physical containment equipment and facilities to reduce the probability of escape of pathogenic microorganisms and viruses from the laboratory environment (27, 38, 47, 61). These include using microbiological safety cabinets, laminar flow biohazard hoods or negative-pressure laboratories, and special traps on drain lines, vacuum lines, or equipment in containment cabinets for any procedure likely to produce aerosols. It is also possible to install autoclaves in the walls of such facilities to make it possible to sterilize material prior to its removal from the room in which a given experiment is done. In addition to the absolute reduction in dissemination of microorganisms, the use of such facilities and equipment has a psychological effect in causing laboratory workers to employ procedures commensurate with the risks associated with the task at hand and to thwart entry of other personnel into the room in which such experiments are done unless direct permission is given by someone in authority. The probability of escape of microorganisms or viral vectors containing foreign DNA would therefore decrease as the physical containment provided by the equipment

and facilities used increased. It should also be noted that the safety barriers associated with any type of physical containment system would be reduced by the absence of appropriate measures to control insects and/or rodents in or around the laboratory.

EMERGENCY OR ACCIDENT PLAN Laboratory accidents are a reality of life and the adoption of an appropriate plan of action in the case of an accidental spill of recombinant DNA or a viral vector or microbe that contains it can markedly reduce the probability of its escape from the laboratory environment. Thus procedures to cope with such accidents should be developed and available for use prior to commencing recombinant DNA molecule experiments.

Factors Affecting the Probability of Survival and Perpetuation of Recombinant DNA

NATURE OF MICROBIAL HOST STRAINS AND CLONING VECTORS In terms of a microbial host strain to be used to introduce recombinant DNA, one can consider, in order of decreasing potential for survival, the use of recently isolated wild-type strains, laboratory-attenuated wild-type strains or their derivatives, and strains whose ability to survive and/or multiply is solely dependent upon uniquely defined laboratory conditions that are unlikely to be encountered in nature, especially in the ecological niches that are normal habitats for the microbial species in question. The survival of the host employed and the possible perpetuation of recombinant DNA by transmission to other microbes also depends on the routes of escape from the laboratory. For spills and aerosolization, resistance of the host to dessication, disinfectants, UV irradiation, etc, is important, and the potential of the host to form cysts, spores, conidia, etc, is relevant. For inadvertent release by disposal down the drain, one must consider the host's survival potential in an aqueous (polluted) environment. Since contamination of personnel is the most likely means of escape for the microbial host in question, the likelihood of its survival on clothing, on skin, in the pharynx, in the oral cavity, during passage through the stomach and the remainder of the gastrointestinal tract, and ultimately in the sewers and during the local sewage treatment processes must be considered. In all these instances, we must examine the probability of the host surviving long enough by any of these potential means of escape to find a suitable environment, either for its multiplication or where other robust microbes present might act as recipients for receipt of the recombinant DNA.

In terms of plasmid-cloning vectors, the use of nonconjugative¹ plasmids affords more containment than use of conjugative¹ plasmids, and nonconjugative plasmids that are poorly mobilizable¹ by conjugative plasmids that might be transferred from microbes in nature to the host containing recombinant DNA are better than nonconjugative plasmids that are readily mobilizable. If the plasmid-cloning vector is

¹A nonconjugative plasmid is unable to promote its own transfer by conjugation, whereas a conjugative plasmid can promote either conjugational transfer of itself or, through a process termed mobilization, the transfer of chromosomal and/or nonconjugative plasmid DNA (63).

GENETIC MANIPULATION OF MICROORGANISMS 515

genetically manipulated so that its replication depends on the host strain or so that its transmission to other microbes would be lethal to them, then one need be concerned only with the survival of the host carrying the recombinant DNA. One must also consider whether the plasmid-cloning vector possesses traits that confer a selective advantage on the host strain that would cause it to be more likely to survive.

Similar considerations can be given for the survival of viral cloning vectors, with wild-type vectors more likely to survive and perpetuate cloned DNA segments than those viral vectors that depend solely on a laboratory-adapted strain for multiplication. If the viral vector can establish a carrier state in the host microorganism, one must consider the survival potential of the host by the various routes of escape discussed above. Likewise, if the viral vector does not solely depend on the host strain for its replication, then one must consider the virus's potential to survive by any route of escape until it encounters a sensitive permissive host cell.

ECOLOGICAL NICHES OCCUPIABLE BY ORIGINAL AND CHIMERIC HOSTS

The probability of survival of escaped chimeric microorganisms is also affected by the natural ecological niches of the host microorganism, the contributions of the foreign DNA to alterations in the potentially occupiable ecological niches, and also the distance between the occupiable ecological niches and the laboratory setting in which the experiments are performed. For viral cloning vectors, a principal factor governing survival and perpetuation of escaped cloned DNA is the proximity of the laboratory environment to the ecological niches in which potential sensitive permissive hosts for the viral vector might exist. In these regards, one must also consider the alterations of the biosphere by current civilization that permit given microbial types to exist in new ecological niches. For example, *E. coli* and other enteric microorganisms that have probably always been prevalent in the intestinal tracts of warm-blooded animals are now found to occur in cold-blooded animals, and higher invertebrates, and certainly in sewers, streams, rivers, lakes, and estuarine waters, as well as on land in both urban and agricultural environments (20). Thus the use of *E. coli* and other enterics for the formation of chimeric microorganisms places normal ecological niches for the organism in or near the laboratory setting. It should therefore be evident that the use of a microbial host that finds humans to be a suitable ecological niche provides the greatest likelihood of successful escape, whereas microorganisms or chimeric microorganisms that find soil and/or plants as normal ecological niches have a somewhat reduced probability of being carried from the laboratory environment to their normal ecological habitat. The use of marine microorganisms as hosts for recombinant DNA would only seem to pose a hazard if the laboratory were located adjacent to a marine environment, and similarly the use of microbes that might occupy a very restricted ecological niche (such as hot springs) would be least likely to ever get from the laboratory environment to their normal ecological habitat.

TRANSMISSIBILITY It is now realized that gene flow among microorganisms in nature is far more common than previously thought; thus one must bear in mind

the likelihood of a cloned DNA fragment being transferred from a viral-cloning vector or chimeric microorganism to some other vector or microorganism in nature or in the same laboratory that has a higher probability of survival. Most, if not all, species within all procaryotic genera are lysogenic for one or more temperate phages (1, 7, 34). Since most temperate phages can probably act either as generalized or specialized transducing phages, it is likely that transduction in nature can lead to transmission of genetic information from one procaryotic microorganism to another in the same or a closely related species. The frequency with which this occurs is likely to be low, although it has only been measured in one instance (62). The probability for transductional transmission of recombinant DNA from the host microbe to other microorganisms in nature can be reduced, if not abolished, by using a nonlysogenic host that is also resistant to transducing phages that are known to infect the host strain. This may be easier said than done, however, because cell surface changes that lead to resistance to known transducing phages are likely to result in sensitivity to other phages that have not been studied and that may be capable of transduction.

Plasmids are now known to occur in at least 40 bacterial genera (31, 46, 55, 56, 68), and it is probable that such plasmids will be found in other bacterial genera as they become better studied. Although conjugation, which is promoted by conjugative plasmids, is best known among gram-negative bacterial genera, the occurrence of some type of conjugational gene transfer in mycobacteria, *Nocardia*, *Actinomyces*, and *Streptomyces* and the newly discovered presence of conjugative plasmids in streptococci (43) causes me to believe that conjugative plasmids are likely to be present in a diversity of other gram-positive genera as well. Wild-type conjugative plasmids, as found in nature, are usually repressed with respect to expression of the donor phenotype, such that only a small percentage of donor cells are capable of plasmid transfer at any given time (31). They are therefore transferred to recipient cells of the same strain at frequencies of 10^{-3} or lower per donor cell. These frequencies may be further reduced by entry exclusion or incompatibility (63) if the recipient also possesses a homologous resident plasmid, by restriction (5), by differences in the cell surface which inhibit the formation of mating unions (22, 30, 57, 67, 75), and by differences in the ability of given strains to establish or maintain a plasmid once transferred into their cytoplasm. These barriers may reduce conjugative plasmid transfer by another several orders of magnitude or abolish it altogether.

Although conjugational gene transfer in *E. coli* occurs at equal frequencies under aerobic and anaerobic environments (77), there are other environmental factors which may act to further reduce conjugational plasmid transfer. I have found that the yield of plasmid transconjugants decreases by the square of the dilution of the parental densities below 10^8 cells/ml, thus transconjugants inheriting plasmids are reduced by three to four orders of magnitude in reducing mating titers from 10^8 to 10^6 cells per ml, the latter being the usual titers for *E. coli* in the intestinal tracts of humans and animals (20). Temperature also exerts an influence on conjugational plasmid transfer, with the optimum for F, I, and N group plasmids commonly found in enterics being near 37°C , with essentially no transfer below 28°C since donor cells are unable to express the donor phenotype at these lower temperatures (10, 55). Thus transfer of these plasmid types in nature probably occurs in intestinal tracts

GENETIC MANIPULATION OF MICROORGANISMS 517

and in fermenting excreta (i.e. manure piles, sewage treatment plants, etc) and not in streams, rivers, or lakes. Conjugative plasmid types commonly found in soil bacteria or in fish pathogens, for example, are likely to be transferred at higher efficiencies at temperatures in the 20° to 30°C degree range, however, so it is probably premature to generalize from the results with the well-studied F, I, and N plasmids found in *E. coli* and its relatives. Conjugational plasmid transfer seems also to be restricted to pH values of 6 to 8.5 (55), requires actively growing cells (23–25), and is diminished in environments of lower nutritive value for the metabolizing bacteria. The probability for transfer of nonconjugative plasmids is indeed very low because of the requirement that a cell harboring such a plasmid must first acquire a conjugative plasmid that is able to mobilize the nonconjugative plasmid for transfer to a suitable recipient. Frequencies of nonconjugative plasmid transfer by such triparental matings may occur under ideal laboratory conditions at frequencies ranging from 10^{-4} per primary donor cell to an unmeasurable quantity, depending on the particular plasmids employed. Thus conjugational plasmid transfer in nature, although obviously occurring, is difficult to quantify at measurable rates or frequencies.

Transformation as a means of gene transfer is known to occur in streptococci, bacilli, *Haemophilus*, and *Neisseria*, and in some instances this has been observed to occur in animal hosts (33). Successful transformation of *E. coli* with viral, plasmid, and chromosomal DNA requires prolonged incubation at 0°C in the presence of CaCl_2 , and then a heat shock to facilitate DNA uptake (17, 21). Such transformation occurs at very low frequencies, even under these unusual conditions (which are probably not encountered in nature). It is thus unlikely that transformation would be a very significant contributing factor for gene flow in nature for most gram-negative microorganisms.

When inheritance of genetic information transferred by virus infection, transformation, transduction, or conjugation requires genetic recombination, the frequency generally depends on similarities in guanine-cytosine contents and nucleotide sequence homologies between donor and recipient DNA. In this regard, however, one must also be aware of the prevalence of insertion sequences (40, 41, 65) and the newly discovered translocatable sequences (36) discovered in enteric microorganisms, which seem to permit, at low frequency, the movement of a DNA segment from one genetic element to another that is independent of any necessary nucleotide sequence homology. Whether such sequences are prevalent in other organisms in nature or are present on cloned DNA fragments remains to be seen. If so, such sequences could facilitate the formation of unusual recombinants by using one or more of the naturally occurring mechanisms for gene flow among microorganisms in nature.

If lower eucaryotic microorganisms are employed as hosts for cloning recombinant DNA molecules, it must be realized that many of these exhibit a classical sexual type of reproduction as well as various types of parasexual gene transfer mechanisms. In either event, the possibility exists not only for transfer of nuclear genetic material but also for cytoplasmic genetic material. These systems, however, are generally restricted to genetic exchange only within a species and then only when mating type factors are complementary.

SELECTIVE ADVANTAGES OR DISADVANTAGES CONFERRED BY FOREIGN DNA In addition to the effects cloned foreign DNA might have on the ecology of the chimeric microorganism and/or the transmissibility of cloning vectors, its presence may alter the survival of cloning vectors or hosts that possess it. The long-term survival and perpetuation of an escaped recombinant DNA molecule necessitates that the foreign DNA confer either some selective advantage or at least no selective disadvantage to viral vectors or chimeric microorganisms that possess it. Probably the replication of the additional foreign DNA in the absence of any benefit to do so is a selective disadvantage. Similarly, the synthesis of gene products specified by the recombinant DNA, whether functional or non-functional, and which bring no benefit to the viral vector or host microorganism imposes an even greater burden. It is likely that in most instances the presence of cloned foreign DNA will be detrimental to the competitive survival of cells and vectors that possess it.

The only available information on this point comes from a series of experiments performed by R. Davis (personal communication). He has taken random DNA segments from yeast, *Klebsiella*, and *Dictyostelium*, cloned them on a λ vector, and then analyzed the frequency distributions of different sizes of foreign DNA segments as a function of the number of repeated λ infection cycles. Although there was initially a large assortment of random fragments, after 30 infective cycles only one or two specific fragments gained ascendancy in each case. When λ vectors containing these surviving fragments were studied in mixed infection with the λ vector lacking a foreign DNA sequence, the λ vector lacking the foreign DNA always became the predominant species. Thus in no case did a foreign DNA segment confer a selective advantage to the λ cloning vector.

Factors Affecting Probable Consequences of Escape and Survival of Recombinant DNA Molecules

The overriding consideration concerns the degree with which the trait specified by the recombinant DNA is compatible with traits possessed by the host microorganism or by other microorganisms that can act as recipients of genetic information from the host microbe, since these relate to the manifestation of a phenotype leading or not leading to the displacement of other organisms from their normal ecological niches. This overall consideration can be approached by first assessing whether or not the cloned DNA information is expressed in the primary host or in hosts that are capable of receiving genetic information from the primary host. As stated above, expression of eucaryotic genetic information in procaryotic organisms is not likely to be universally observed without additional genetic manipulation to permit faithful transcription and translation of eucaryotic genes. The reciprocal situation is equally true, but probably it is not a problem with regard to expression of eucaryotic genetic information in eucaryotic microbial hosts or of procaryotic genetic information in procaryotic microbial hosts. I anticipate, however, that difficulties in expression of foreign DNA will ultimately be eliminated and thus in the following discussion I assume that foreign gene expression is likely to occur.

With respect to the compatibility of an expressed foreign-DNA-specified trait in a microbial host or in other microorganisms with which the primary host can

GENETIC MANIPULATION OF MICROORGANISMS 519

exchange genetic information, there is a range of possibilities. For example, endowing *E. coli* with the ability to synthesize keratin, if that were possible, would be unlikely to lead to hairy bacteria, whereas the endowment of *E. coli* with the capability to carry out photosynthesis, because of transmission of such genetic information to other gram-negative soil-inhabiting microorganisms, might lead to the potential for a green earth. Also related to the expression of foreign genetic information is the ability of the primary host microorganism or of other microorganisms with which it can exchange genetic information to produce a gene product that can be released into the biosphere by occasional lysis of cells or by active secretion, in such a way that it has an adverse effect on the physiology, metabolism, or livelihood of other organisms that might come in contact with the substance. In these regards, then, we can look at competitive displacement of other organisms from their normal ecological niches as resulting from competition for essential foodstuffs, the production of products that are deleterious to other organisms, and/or the exhibition of traits of general pathogenicity, virulence, and invasiveness.

In considering the potential for surviving recombinant DNA, whether in the original or in a secondary host, to cause displacement of other organisms from their normal ecological niches, one must also realize that differences in susceptibility of other organisms may depend upon their age. For example, newborn mammals have an immature intestinal mucosa that permits the transport of macromolecules; thus newborns could be susceptible to gene products produced by chimeric microorganisms that would not affect adults of the species. In a similar vein, one must also consider food chains in which an organism that might be displaced by chimeric microorganisms might initially appear to have no relevance to us, but which might play an important role as predator of another organism that, if unchecked, might become a serious pest or even a vector of an infectious disease agent active against other organisms in the biosphere. Similarly, the displaced organism might itself be prey and/or an essential foodstuff for another organism that might be of more direct importance to humankind. These types of subtle effects are difficult to predict and give cause for concern about the inadvertent escape and survival of recombinant DNA molecules.

*Estimating the Probability of Manifestation of Adverse Consequences
by Recombinant DNA Molecules*

ESCAPE In its proposed guidelines for research involving recombinant DNA molecules, the National Institutes of Health has stipulated four categories of physical containment, with their associated procedural precautions. These categories, designated P1, P2, P3, and P4, are to be employed commensurate with the potential risks associated with any given experiment, depending on whether these risks are minimal or nonexistent, low, moderate, or high. If all materials are sterilized or disinfected prior to disposal (a requirement for all four categories of containment), then this should all but eliminate the possibility that chimeric microorganisms or viral vectors containing foreign DNA could escape down the sink. If there is no mouth pipetting (advisable for P1 containment and a requirement for the P2, P3,

and P4 containment categories), then accidental ingestion of viral vectors or hosts containing foreign DNA should also be a remote possibility. However, many common laboratory procedures, as well as accidental spills, produce aerosols, and this production of aerosols is most likely to lead to contamination of investigators and the laboratory environment (28). Table 1 provides data on the doses of viruses or microorganisms that are likely to be inhaled because of the production of aerosols during several common laboratory operations performed by both proper and improper techniques. Such operations as inoculating cultures, pipetting, mixing (vortexing), centrifugation, and so on also produce aerosols (28). It should be noted that the contamination of investigators and the laboratory environment due to aerosols make it all the more important that investigators wear appropriate protective clothing which should not be worn outside the laboratory, refrain from eating, drinking, and smoking in the laboratory, since these practices lead to the ingestion and/or inhalation of aerosolized materials, wash their hands with disinfectant soap upon completion of procedures (disposable gloves which would be autoclaved are required for some experiments using P3 and P4 containment), and disinfect the work area upon completion of procedures.

It is thus apparent that, even when acceptable practices are used in the type of facilities required for a given experiment, some virus vectors or host organisms will escape from the laboratory environment. I estimate that the numbers of escaped viral vectors or host microorganisms should be reduced about 1000-fold for each increasing level of containment starting with P1. Estimating the numbers of virus vectors or hosts that might escape per day per investigator for a properly operated P1 facility is difficult, but let me guess at a value of 10^5 . If we then assume 1000 investigators working on 300 days per year, then 3×10^{10} escape from P1 facilities, 3×10^7 from P2 facilities, and 3×10^4 from P3 facilities. These routes of escape are by contamination of the investigator (on clothing or by breathing or ingestion), by discharge through the ventilation system, by removal of floor sweepings, and by accidental disposal down the drain.

Some members of the scientific community have argued that P1 and P2 containment afford no barriers to escape of chimeric microorganisms and therefore have suggested that P3 containment be the minimum requirement for recombinant DNA molecule research. They base their arguments, in part, on the observation that many microbiologists do not use such facilities and procedures properly, and they fear that many nonmicrobiologists who use such facilities and procedures will not initially know how to use them properly. Although I concede that this may be so and acknowledge that the estimates made above for the escape of virus vectors and microbial hosts would be increased considerably if the appropriate facilities and procedures are improperly used, I believe that one cannot legitimately redefine standard microbiological practices that are based on years of experience. For example, most clinical microbiology laboratories use P1 or P2 facilities and procedures and P2 facilities are sufficient for research with *Bacillus anthracis*, *Clostridium botulinum*, *Corynebacterium diphtheriae*, *Salmonella typhi*, *Shigella dysenteriae*, *Streptococcus pneumoniae*, *Vibrio cholera*, Adenoviruses, Herpes virus hominus, Simian virus 40, measles virus, and so on (14). It thus would be inappropriate to

Table 1 Estimated inhalation dosages of viruses or microorganisms during laboratory operations^a

| Procedure | Person 1 | | | | Person 2 | | | | Person 3 | | | |
|--|----------------------|----------------------|-----------------------|--|------------------------|------------------------|-------------------------|--|----------------------|-----------------------|-----------------------|--|
| | P1 | P3 | P4 | | P1 | P3 | P4 | | P1 | P3 | P4 | |
| Flame sterilization of bacteriological loop ^b | | | | | | | | | | | | |
| Good procedure | 1 × 10 ⁻³ | 1 × 10 ⁻⁸ | 1 × 10 ⁻¹² | | 1 × 10 ⁻⁶ | 1 × 10 ⁻¹¹ | 1 × 10 ⁻¹⁵ | | 1 × 10 ⁻⁸ | 1 × 10 ⁻¹⁴ | 1 × 10 ⁻¹⁸ | |
| Poor procedure | 1 × 10 ⁻¹ | 1 × 10 ⁻⁶ | 1 × 10 ⁻¹⁰ | | 1 × 10 ⁻⁴ | 1 × 10 ⁻⁹ | 1 × 10 ⁻¹³ | | 1 × 10 ⁻⁶ | 1 × 10 ⁻¹² | 1 × 10 ⁻¹⁶ | |
| Sonication ^c | | | | | | | | | | | | |
| Good procedure | 3 × 10 ⁰ | 3 × 10 ⁻⁵ | 3 × 10 ⁻⁹ | | 1.5 × 10 ⁻² | 1.5 × 10 ⁻⁷ | 1.5 × 10 ⁻¹¹ | | 1 × 10 ⁻⁴ | 1 × 10 ⁻¹⁰ | 1 × 10 ⁻¹⁴ | |
| Poor procedure | 3 × 10 ³ | 3 × 10 ⁻² | 3 × 10 ⁻⁶ | | 1.5 × 10 ¹ | 1.5 × 10 ⁻⁴ | 1.5 × 10 ⁻⁸ | | 1 × 10 ⁻¹ | 1 × 10 ⁻⁷ | 1 × 10 ⁻¹¹ | |
| Blending ^d | | | | | | | | | | | | |
| Good procedure | 4 × 10 ⁻² | 4 × 10 ⁻⁷ | 4 × 10 ⁻¹¹ | | 2 × 10 ⁻⁴ | 2 × 10 ⁻⁹ | 2 × 10 ⁻¹³ | | 1 × 10 ⁻⁶ | 1 × 10 ⁻¹² | 1 × 10 ⁻¹⁶ | |
| Poor procedure | 8 × 10 ³ | 8 × 10 ⁻² | 8 × 10 ⁻⁶ | | 4 × 10 ¹ | 4 × 10 ⁻⁴ | 4 × 10 ⁻⁸ | | 2 × 10 ⁻¹ | 2 × 10 ⁻⁷ | 2 × 10 ⁻¹¹ | |
| Drop and break flask ^e | 2 × 10 ² | 2 × 10 ⁻³ | 2 × 10 ⁻⁷ | | 7 × 10 ⁻¹ | 7 × 10 ⁻⁶ | 7 × 10 ⁻¹⁰ | | 1 × 10 ⁻³ | 1 × 10 ⁻⁹ | 1 × 10 ⁻¹³ | |

^a Laboratory is 10 × 20 × 10 ft with a ventilation rate of 10 changes of air per hr. All operations are assumed to be with virus and/or microorganism suspensions of 10⁹/ml. It is also assumed that all aerosols produce particles in the 0.5 to 5.0 μm diameter size range that penetrate and are retained by the lungs. Person 1 is doing the operation. Person 2 is in the same laboratory room. Person 3 is in the corridor just outside the closed door to the laboratory. P1 refers to performance of the operation on a laboratory bench. P3 refers to performance of the operation in a biohazard vertical laminar flow hood. P4 refers to performance of the operation in a sealed vertical laminar flow glove box. Data supplied by W. E. Barkley and M. A. Chatigny (personal communication).

^b Using full loop with 1 min exposure time. Good procedure is heating of shank to kill all organisms prior to flaming loop. Poor procedure is introducing loop directly into flame.

^c Sonication for 1 min with 10 min exposure time. Good technique is with minimal aeration and no frothing. Poor technique is with maximum aeration and frothing.

^d Blending for 1 min with 10 min exposure time. Good technique is using a containment safety blender, with removal of blended material through stoppered entry ports by using a syringe or a regular blender, but waiting one hour before removing top and blended material. Poor technique is using a regular blender with removal of top and blended material immediately after blending.

^e An accident with 1 min exposure time.

penalize all those who do know how to use P1 and P2 facilities correctly and procedures and who wish to conduct low-risk experiments by requiring that expensive P3 conditions be the minimum. On the contrary, it should be the ethical responsibility of all investigators conducting recombinant DNA molecule research to ensure that their collaborators and support personnel follow the acceptable procedures for the type of experiment being performed. Peer pressure and review, assumption of liability for adverse consequences due to recombinant DNA molecule research by the principal investigator and the institution, and the possibility that the investigator and/or the institution might be deprived of government research support should ensure this.

SURVIVAL In assessing the probability of survival of recombinant DNA molecules in nature, given that accidental escape has occurred, we suffer from a paucity of factual information. What follows therefore represents my best judgments based on the available data and my own experiences. We must first consider the survival of the host or viral vector which carries foreign DNA and then the potential for survival of the recombinant information if it is transmitted to another organism in nature. Because most current studies are being conducted with *E. coli* K12, I consider it as an example to estimate the probabilities for survival of recombinant DNA. In the guidelines for recombinant DNA molecule research being promulgated by the National Institutes of Health (NIH), several levels of biological containment for *E. coli* K12 host-vector systems are stipulated. The EK1 systems utilize existing *E. coli* K12 strains either with nonconjugative plasmid or λ phage vectors and are supposed to afford a moderate level of containment. The EK2 systems must be genetically modified to provide a high level of containment so that the probability of survival of cloned foreign DNA is 10^{-8} or less, either by survival of the host or by transmission and survival of the cloned DNA in some other host or vector.

E. coli K12 was isolated from a human patient in 1922 but since that time has become adapted to the laboratory environment (6). Although it will survive passage through the intestinal tracts of humans, calves, rats, and mice, it has failed to colonize these species in a limited number of studies (3,4,76; S. Falkow, personal communication; R. Curtiss and co-workers, unpublished). These experiments, however, are of limited value in reaching any valid generalizations because the factors leading to colonization of the intestinal tract are poorly understood. For example, *E. coli* strains isolated from one strain of mice are generally unable to colonize other strains of mice or other rodent species but pass through the intestinal tract in much the same way as does *E. coli* K12. Thus there seems to be some type of mammalian host species or strain specificity that plays a role in colonization. The method of prior growth of the *E. coli* strain also influences the degree of colonization or persistence of the strain in the intestinal flora with anaerobic growth leading to longer persistence than aerobic growth. In the studies with feeding of *E. coli* K12 to humans (3,4, 76), variables such as age and natural diet of the individuals, which should also influence colonization, were not systematically studied. *E. coli* K12 cells surviving passage through the intestinal tract remain viable in feces for several days

GENETIC MANIPULATION OF MICROORGANISMS 523

at least, but little or no information is available on the survival of this strain in sewage or during passage through sewage treatment plants. The plentifulness of *E. coli* strains in sewers, rivers, etc, leads one to believe that this species not only survives for some period of time in these environments, but must also undergo active multiplication. Certainly, unpolluted waters receiving vigorous aeration are not favorable for survival and multiplication of *E. coli* or other enteric microorganisms. Thus the quality of receiving waters, which is generally low throughout the world, is a major factor in the ultimate survival of inadvertently released, chimeric *E. coli* strains. *E. coli* strains are found in the pharynx and oral cavity of healthy individuals on an intermittent basis (Curtiss et al, unpublished), but little is known relative to the likelihood of *E. coli* K12's survival in these ecological niches, or for that matter, its survival on the skin. There is also little information about the survival of *E. coli* K12 during conditions of drying following a laboratory spill, although we have found a 50–90% death rate per day with *E. coli* cells dried under sterile conditions in a growth medium.

Based on these considerations, I estimate that escaped EK1-type *E. coli* K12 cells would have a mean probability of survival of about 10^{-1} per day, although one must consider that multiplication might occur in some favorable environments such as a sewage treatment plant. By using an EK2 *E. coli* K12 host whose survival is solely dependent upon a uniquely defined laboratory environment that is unlikely to be encountered in nature, we have found that the probability of survival is 10^{-8} or less after one day (or less) under conditions that are favorable for growth of normal *E. coli* cells and is about 10^{-1} per day under conditions that do not permit growth of normal *E. coli* (i.e. in water, physiological saline, after drying, etc).

In terms of transmission of genetic information from *E. coli* K12 strains in nature, we can consider in decreasing order of likelihood conjugation, transduction, and transformation. *E. coli* K12 is known to be able to receive or donate plasmid DNA with representative strains from over 20 bacterial genera under optimal laboratory controlled conditions (31, 46, 55, 56, 68). Nevertheless, the frequencies of transmission of plasmids in certain plasmid incompatibility groups borders on the level of undetectability, even under these optimal conditions. Transmission of some plasmids, however, is known to occur in the intestinal tracts of various mammalian and avian species that usually possess an artificial flora because they began as gnotobiotic animals or were treated with antibiotics prior to contamination (4, 31). In normal individuals, the frequency of conjugative plasmid transmission is all but undetectable (3, 4, 31). I would therefore estimate that if an F or I group conjugative plasmid (repressed for expression of the donor phenotype) were used to clone DNA in wild-type *E. coli* K12 cells, the probability for transferring the chimeric plasmid to another microorganism would be no more than 10^{-8} per surviving escaped cell. (This estimate is based on the maximum observable frequency of 10^{-3} per donor cell for F or I group repressed conjugative plasmid transfer under ideal laboratory conditions, which is decreased a further 100-fold because the titers of suitable recipients in nature are about 10^6 /ml rather than 10^8 /ml, and another 1000-fold because of other barriers to plasmid transmission such as restriction, entry exclusion, incompatibility, cell surface differences that block union formation, and/or inability of

recipient cells to establish and/or maintain the donated plasmid.) This probability is reasonably valid for escaped *E. coli* K12 chimeric hosts that are either ingested or disposed of down the drain where they will ultimately end up in a sewage treatment plant, but is unreasonably high for chimeric hosts that escape by aerosolization and end up in a dried state that is not conducive to either finding a suitable partner or conjugation. This probability would also be further decreased if the chimeric conjugative plasmid is in a disabled EK2 host, since a metabolically inactive donor very rapidly loses its ability to conjugate (24). If a nonconjugative plasmid such as pSC101 is employed as the cloning vector, the chimeric microbe must first acquire a conjugative plasmid, which might occur at a probability of no more than 10^{-7} per surviving chimeric host cell. (The slightly higher probability for acquiring a conjugative plasmid is because host strains used for recombinant DNA molecule research are generally restrictionless, thus eliminating this barrier.) Since this rare chimeric host cell that now possesses a conjugative plasmid would have a probability of no more than 10^{-8} per surviving cell of conjugational transmission with some other microorganism and since the probability of pSC101 mobilization by most conjugative plasmids is about 10^{-5} to 10^{-7} per successful conjugational event, the overall probability for transfer of a chimeric pSC101 plasmid would probably not exceed 10^{-20} per surviving cell. If the *E. coli* K12 host used has been genetically manipulated to reduce or eliminate its ability to accept conjugative plasmids, then the probability for conjugational transmission of a chimeric nonconjugative plasmid can be essentially eliminated. This same result can also be achieved if the nonconjugative plasmid either depends for its replication on its host strain or possesses a killer function that leads to the death of other microorganisms into which it is transferred.

The ubiquity of conjugative and nonconjugative R plasmids and other plasmid types specifying surface antigens, colicins, enterotoxins, and so on would suggest, however, that although the frequencies of conjugational plasmid transmission are generally unmeasurable (i.e. occur at frequencies lower than 10^{-8}), they do occur in nature. In this regard, one must remember the immense numbers of bacteria that occupy a given sewage treatment plant or a given polluted river, so that probabilities of occurrence as low as 10^{-20} may not be all that infrequent. Indeed, the total number of *E. coli* cells within the intestinal tracts of the world's entire human population is very nearly equal to 10^{20} .

The probabilities for transmission of chimeric plasmid DNA by transduction to other microorganisms is difficult to estimate. Specialized transducing phages are not likely to be a problem since the probability for integration of a prophage into a plasmid vector is likely to be 10^{-6} or lower, and then upon spontaneous induction (occurring with a probability of about 10^{-4}) only one in 10^6 phage might possess a recombinant DNA fragment. Thus, the probability for getting a specialized transducing phage carrying recombinant DNA would already be 10^{-16} per surviving chimeric *E. coli* K12 cell even before estimating the probability that this surviving cell will encounter a suitable phage in nature and then the probability that the transducing phage produced would survive long enough to encounter a robust sensitive *E. coli* cell to infect and lysogenize. Generalized transducing phages such

as P1 (and its relatives), Mu, D108, and T1 could, upon infection of a surviving *E. coli* K12 cell, pick up the entire chimeric plasmid with probabilities ranging from 10^{-5} to 10^{-9} , depending on the particular phage. Again we must also consider the probabilities of the surviving chimeric *E. coli* K12 cell being infected by such phages and of the transducing phage encountering a sensitive strain that can be transduced. It should also be mentioned that transduction by any of these phages will also be decreased 10- to 1000-fold by the probable occurrence of restriction by the infected sensitive host. Transductional transmission of recombinant DNA can be made an even less likely event by using an EK2 host strain that is resistant to common *E. coli* transducing phages and that is, outside of its carefully controlled laboratory environment, metabolically defective in terms of permitting productive phage infection.

The probability for transmission of chimeric plasmid DNA by transformation should be dependent on the lysis of the *E. coli* K12 host, the survival of the DNA in the absence of degradation and the likelihood that this DNA could be taken up by and survive restriction in suitable microorganisms in nature. Since transformation of gram-negative microbes requires incubation of cells at 0°C for several hours in the presence of CaCl_2 , followed by a heat shock (17)—an unlikely set of conditions to find in nature—and since even this treatment only results in a 10^{-6} probability of transformation in nonrestricting hosts, it seems most improbable that recombinant DNA could be transmitted in this manner. The use of an EK2 *E. coli* host that undergoes damage to its DNA outside of its carefully controlled laboratory environment would further minimize the probability for transmission by transformation in nature.

In terms of viral cloning vectors, most experiments now being done are utilizing bacteriophage λ of *E. coli* K12. Most λ cloning vectors now in use are defective in lysogenization and plasmid formation [each type of occurrence has a probability of 10^{-6} or less (50, 53, 71, 72)], and because of this one must consider survival of the propagating host strain, which could be either a typical EK1 laboratory strain or an EK2 strain that depends on carefully controlled laboratory conditions for its growth and survival. In the former case, the probability of survival per day would be about 10^{-1} , and in the latter case between 10^{-1} and 10^{-8} , depending on the absence or presence of conditions permitting rapid death of the EK2 host. We would thus have combined probabilities during the first day for formation and survival of λ cloning vectors that contain foreign DNA as either a plasmid or prophage of between 10^{-7} and 10^{-14} , although the EK2 host could not survive in nature and would have to transmit the foreign DNA to some other robust bacterial strain. The use of chloroform, which is customary when making phage lysates (1), would all but eliminate these routes of escape and survival, however. The probability of survival of foreign DNA carried on escaped λ particles depends on the probabilities of λ survival and its encountering a suitable sensitive host in nature. Wild-type λ is relatively sensitive to dessication and to the acidity of the stomach (W. Szybalski, personal communication) and λ -sensitive *E. coli* strains are relatively rare in nature (R. Davis, personal communication), probably no more frequent than one per 1000. Thus the probability of an escape λ surviving and encountering such a host might

be between 10^{-9} and 10^{-12} during the first day after escape. Because most such hosts would restrict λ , this decreases these probabilities another 10- to 1000-fold. If the λ cloning vector is dependent on its original host for its propagation, however, and also is highly virulent and unable to establish lysogeny, then the probability for transmission of carried foreign DNA depends on the likelihood of an encountered λ -sensitive *E. coli* possessing a lambdoid phage as a prophage and of recombination yielding a wild-type lambdoid phage possessing the foreign DNA. Even though lambdoid phages are reasonably common in wild-type *E. coli* strains (44, 64), this required sequence of events should reduce the overall probability of perpetuation of the foreign DNA cloned on λ to less than 10^{-20} .

Another factor that reduces the probability of recombinant DNA survival is the recent finding that foreign DNA sequences cloned on bacteriophage λ are lost in competition with wild-type λ during multiple replication cycles (R. Davis, personal communication). Of course, certain cloned DNA fragments may either increase, or at least not adversely affect, the survival potential of cells or vectors that possess them, but these instances seem to be rare ($P \approx 10^{-3}$) compared to the total types of cloned DNA fragments that arise from a shotgun type of experiment in which the total genome of the donor is fragmented and introduced randomly into the cloning vector.

ADVERSE CONSEQUENCES In considering adverse consequences due to expression of foreign DNA, I assume that expression can be achieved. If one then starts with random fragments of a foreign genome, I estimate that no more than one in 1000 such fragments possess genetic information that would be deleterious to some other organism in the biosphere if expressed by a chimeric microorganism in the same ecological niche. One could also reduce this probability for expression of adverse consequences another 10-fold at least, because the deleterious gene product is unlikely to be secreted or excreted in a biologically active form or then to be taken up by the target organism where it can exert its adverse effect. Thus the overall probability that an escaped surviving random foreign DNA segment would manifest expression of adverse consequences for some organism in the biosphere is 10^{-4} or less. It should be obvious, however, that this value should be modified depending on the source, purity, and known potential for expression of harmless vs harmful gene products of the cloned foreign DNA.

One other cautionary comment should be raised about the potential for manifestation of adverse consequences associated with recombinant DNA molecule research. Namely, will any of the supposedly beneficial applications of the technology create a biohazardous condition that would be more detrimental than the original problem being corrected? A meaningful answer to this question is probably premature, but it would certainly be worthwhile to give careful thought to this issue before advancing too far in using recombinant DNA molecule techniques to solve the myriad of problems now confronting our society.

SUMMATION In the foregoing treatment, I attempt to estimate all factors that would or could contribute to the display of a biohazardous situation by using, as

GENETIC MANIPULATION OF MICROORGANISMS 527

examples, the cloning of random DNA sequences of a eucaryotic organism with *E. coli* K12 hosts and vectors. For EK1 plasmid-host systems, the limiting factor is the probability of survival of the host, and thus one can discount the lower probabilities for transmission of cloned foreign DNA in estimating the likelihood of biohazardous situations. With λ vector EK1 systems, however, one should consider both the survival of the host if lysogenization and/or plasmid formation can occur, as well as the survival and perpetuation of the λ vector. For EK2 plasmid-host systems in which the host theoretically cannot survive in nature, the probability of perpetuation of cloned foreign DNA is the product of the host survival and the transmissibility probabilities. The use of EK2 hosts with EK2 phage or plasmid vectors that either are dependent on the EK2 host or that kill other microorganisms into which they are introduced all but eliminate the probability for survival and perpetuation of the cloned DNA. In these instances the probabilities need to be multiplied times the total number of escaped chimeric microorganisms or viral vectors times the probability that the cloned DNA would not confer a selective disadvantage on hosts or viral vectors times the probability for displaying adverse consequences.

Although these total estimated probabilities for an occurrence of a biohazard are exceedingly small, it should be reiterated that noncompliance with use of standard required practices for physical and biological containment makes these low estimates meaningless. For example, contamination of an EK2 host culture at the time of cloning foreign DNA could yield robust chimeric microorganisms with excellent chances for survival and perpetuation of cloned DNA.

The same considerations for estimating biohazard probabilities can be made for cloning foreign DNA in other microorganisms such as *Bacillus*, *Pseudomonas*, yeast, and others. The favored routes for escape and ecological niches for survival and the organisms at risk in the biosphere are different than for *E. coli*, but the ultimate long-term effects would be similar.

INTENTIONAL RELEASE OF GENETICALLY ALTERED MICROORGANISMS

For Beneficial Purposes

Some foreseeable applications of the ability to clone foreign DNA in bacteria include use of these chimeric microbes for beneficial purposes to clean up environmental pollutants, to eliminate specific insect pests, or even to produce antibodies specific against toxic microbial products, such as the cholera enterotoxin as a means to cope with cholera infections. These beneficial purposes may be laudatory, although I consider such release any time in the near future to be potentially quite hazardous and therefore totally inadvisable. For one thing, a substantial amount of information would be needed to verify that the chimeric microorganism exhibited all the desired attributes and none of the potential undesirable attributes. Furthermore, the release of such a microorganism in one country is a de facto release in all other countries, irrespective of whether such other countries approve or not. Therefore, some inter-

national authority needs to be established to govern or regulate such beneficial uses of chimeric microbes.

For Nonbeneficial Purposes

Although reprehensible and agonizing to contemplate, the technologies for constructing chimeric microorganisms could be utilized to develop biological weapons. Despite the recent signing of treaties to destroy all biological weapons and cease further development of such agents, recent events have undermined the faith of the individual in the credibility of the governments of the world. One must therefore exhibit some degree of apprehension about the potential misuse of these technologies. I would hope that anyone or any government contemplating such use would recognize that the consequences would be just as adverse for them as for any other group of individuals.

COPING WITH POTENTIAL BIOHAZARDS

Shortly after the publication of the Berg et al (18) letter calling attention to the potential biohazards associated with recombinant DNA molecule research, I began to consider the possibility that one might genetically alter *E. coli* as a host for recombinant DNA research to make it safer, a goal in keeping with our then-current intent to make *E. coli* more useful for recombinant DNA molecule research. This idea was communicated to a group of microbial geneticists and plasmid researchers who had been asked to draft a proposal for the conduct of recombinant DNA molecule experiments with procaryotic organisms for submission to the Asilomar Conference on Recombinant DNA Molecules. These concepts were discussed, expanded, and refined by this group and presented at this meeting. The concept of genetic manipulation of vectors and hosts to provide biological containment for recombinant DNA molecule research was persuasively advocated by S. Brenner at the Asilomar meeting and resulted in the adoption of this added means to contend with potential biohazards (9).

Stated simply, as it would apply for the use of any microbial system for recombinant DNA molecule research, the host strain should be uniquely adapted to a carefully controlled laboratory environment so that it can neither survive nor transmit genetic information to other organisms in any environment in which it might naturally find itself. This obviously means that the host strain should not be able to survive in its normal ecological niches if these should be present in the laboratory setting or connected or adjacent to it; for example, intestinal microorganisms should not be able to colonize or survive in the intestine, organisms that normally reside on the skin should not be able to survive on the skin, organisms that normally live in soil or on plants should not be able to live in soil or on plants, and so on. In terms of transmissibility of genetic information, it is best to have a plasmid or viral cloning vector that is solely dependent for its replication on the genetically altered host and even to endow it with the potential to kill any other microorganism into which such DNA might be introduced. In so genetically modifying vectors and hosts, one should employ deletion mutations whenever possible and even double mutations

GENETIC MANIPULATION OF MICROORGANISMS 529

with the mutant alleles at widely spaced positions on the genome to diminish the likelihood of reversion of the trait or loss by genetic exchange and/or recombination with other organisms or vectors in nature.

In the year since the Asilomar Conference, our laboratory group has endeavored to construct a number of *E. coli* K12 strains that would provide these fail-safe features. At the commencement of these endeavors, the task seemed straightforward; however, as we soon learned, appropriate deletion mutations did not exist and *E. coli* had a will as well as a means to survive under adverse conditions when it should have dropped dead. It has thus required a total of 13 genetic manipulations to obtain a strain that should meet the requirements for an EK2 host as specified in the current proposed NIH Guidelines for Recombinant DNA Molecule Research. This strain cannot synthesize the murein layer of its cell wall, except under carefully controlled laboratory conditions because it has two mutations at different map positions; is unable to synthesize the mucopolysaccharide colanic acid that can facilitate the ability of *E. coli* to survive in the absence of the murein layer of the cell wall; is sensitive to bile salts and detergents, undergoes thymineless death, and degrades its DNA in environments lacking thymine or thymidine; is resistant to a number of *E. coli*-transducing phages; is very sensitive to UV irradiation; is unable to act as a recipient in matings with donors possessing conjugative plasmids in five of the groups commonly or occasionally found in enteric microorganisms; and, finally, is defective as a recipient in matings with donors possessing other conjugative plasmid types. This strain still has a finite low but unmeasurable probability for transmitting recombinant DNA molecules to other organisms in nature, and we are therefore continuing our studies to design a completely fail-safe bacterial host—if that is at all possible.

Our recent experience in this endeavor leads me to believe that it would be possible to genetically alter other microorganisms and also cloning vectors in such a way as to all but eliminate any and all of the potential biohazards that might be associated with recombinant DNA molecule research. This task would not be easy, however, and would require considerably greater amounts of information than are currently available on the genetics, metabolic capabilities, and ecological potentials of microorganisms that might be deemed worthy for use as hosts for recombinant DNA molecule research. Until such time, it is likely that *E. coli* K12 will continue to be the workhorse of modern molecular genetics. It should be readily admitted that many have criticized the use of *E. coli* for this research because of its intimate association with humans, its known potential for exchange of genetic information with strains of other gram-negative genera, and the potential of other non-laboratory-adapted strains of *E. coli* to exhibit considerable virulence, pathogenicity, and communicability. Although I agree that these known attributes of *E. coli* make it less than totally acceptable for use in these experiments, these problems are well known and the means to effectively eliminate these problems have now been developed. To choose another microbial host about which less is known under the expectation that such similar problems do not exist seems irresponsible to me. It is far better to know a sufficient amount of information in order to recognize and then eliminate all problems than to assume they do not exist and proceed on that basis.

530 CURTISS

DATA NEEDED TO EVALUATE POTENTIAL BIOHAZARDS OF CHIMERIC MICROORGANISMS

Although information will eventually be learned during recombinant DNA molecule research that will permit assessments of potential biohazards, I believe that a more dedicated effort needs to be made by scientists engaged in this research to obtain as much of this information as soon as possible after construction of chimeric microorganisms or virus vectors. For example, experiments to determine whether the cloned DNA affects the survival of the host under permissive and nonpermissive conditions or to determine the transmissibility of the plasmid vector should have a high priority. Competition experiments can also be performed to determine whether a given recombinant DNA molecule causes a cell to grow slower or faster than a host just carrying the cloning vector or whether it causes a cell to display a selective advantage or disadvantage over other cells of the same or different species that lack the recombinant DNA. Studies can also be performed to determine whether the recombinant DNA is expressed and if so whether the products synthesized are deleterious to any organism that might share or provide an ecological niche for the chimeric microorganism or other organisms with which it can potentially exchange genetic information. Similar experiments can also be performed for viral vectors containing foreign DNA. In all these types of studies, one must consider the source of foreign DNA, the traits potentially specified, and the properties of the host microorganism or viral vector in evaluating which organisms might be at potential risk and which therefore should be studied for adverse effects caused by the host microorganism or viral vectors carrying foreign DNA. In association with measurements as to whether the presence of the cloned DNA alters the frequency with which the recombinant DNA can be transmitted to other microorganisms, it would also be desirable to gradually obtain additional data on the prevalence of microorganisms in nature which display sensitivity to viruses being used as cloning vectors, which possess nonsense suppressors (since these might permit replication of cloning vectors that are dependent on such suppressors), and which lack restriction systems.

CONCLUDING REMARKS

Recombinant DNA molecule research will undoubtedly contribute information that will revolutionize biological science and ultimately society in the years to come. The conduct of this research is associated with numerous potential biohazards that are difficult, if not impossible, to assess totally at this time because of a lack of needed information. Nevertheless, the application of established procedures and physical containment facilities for the conduct of this research, along with the use of disarmed hosts and cloning vectors that are unable to permit the survival and perpetuation of cloned DNA sequences, should all but eliminate any adverse consequences of this research. In view of the lack of information about the basic ecology of microorganisms and their interactions with other species and about the deleterious attributes that might be associated with recombinant DNA, it is urged that scientists engaged in this research endeavor to learn as much about these potential biohazards

GENETIC MANIPULATION OF MICROORGANISMS 531

as is possible, so that future assessments of hazards can be far less speculative. Although it now seems likely that members of the scientific community who plan to do this research have accepted a set of guidelines that maximize benefits and minimize hazards, it is far less certain that similar guidelines have been or will be established to guide the commercial applications of this technology or the misuse of the technology for the preparation of military weapons. For these reasons I remain apprehensive.

ACKNOWLEDGMENT

I thank my numerous scientific colleagues who have entered freely into discussion and debate during the past 18 months about all aspects of recombinant DNA molecule research and who have thus had a major impact on my thinking. I owe a special debt of gratitude to R. C. Clowes, S. N. Cohen, S. Falkow, and R. P. Novick, to those who communicated unpublished information, and especially to the members of my current laboratory group, D. Pereira, J. E. Clark, S. Hull, J. C. Hsu, R. Goldschmidt, R. Moody, L. Maturin, and M. Inoue, who have had an especially pronounced influence on my opinions, attitudes, and thinking and who have labored diligently to disarm *E. coli*. Research is supported by grants from the National Science Foundation (GB-37546) and the National Institutes of Health (DE-02670 and AI-11456).

Literature Cited

1. Adams, M. H. 1959. *Bacteriophages*. New York: Interscience. 592 pp.
2. Advisory Board for the Research Councils. 1975. *Report of a Working Party on the Experimental Manipulation of the Genetic Composition of Micro-Organisms*. Presented to Parliament by the Secretary of State for Education and Science by Command of Her Majesty. January 1975. London: Her Majesty's Stationery Office, 1975.
3. Anderson, E. S. 1975. *Nature* 255: 502-4
4. Anderson, J. D., Gillespie, W. A., Richmond, M. H. 1974. *J. Med. Microbiol.* 6:461-73
5. Arber, W., Linn, S. 1969. *Ann. Rev. Biochem.* 38:467-99
6. Bachmann, B. J. 1972. *Bacteriol. Rev.* 36:525-57
7. Barksdale, L., Arden, S. B. 1974. *Ann. Rev. Microbiol.* 28:265-99
8. Berg, P., Baltimore, D., Boyer, H. W., Cohen, S. N., Davis, R. W., Hogness, D. S., Nathans, D., Roblin, R. O., Watson, J. D., Weissman, S., Zinder, N. D. 1974. *Science* 185:303; *Proc. Natl. Acad. Sci. USA* 71:2593-94; *Nature* 250:175
9. Berg, P., Baltimore, D., Brenner, S., Roblin, R. O., Singer, M. F. 1975. *Science* 188:991; *Nature* 225:442; *Proc. Natl. Acad. Sci. USA* 72:1981
10. Bonhoeffer, F. 1966. *Z. Vererbungsl.* 98:141-49
11. Brown, D. D., Stern, R. 1974. *Ann. Rev. Biochem.* 28:667-93
12. Chang, A. C. Y., Cohen, S. N. 1974. *Proc. Natl. Acad. Sci. USA* 71:1030-34
13. Chang, A. C. Y., Lansman, R. A., Clayton, D. A., Cohen, S. N. 1975. *Cell* 6:231-44
14. *Classification of Etiologic Agents on the Basis of Hazard*, July 1974. U.S. Dept. HEW, Public Health Serv. Atlanta: Ctr. Disease Control, Office Biosafety. 4th ed.
15. Cohen, S. N. 1975. *Sci. Am.* 233:24-33
16. Cohen, S. N., Chang, A. C. Y. 1974. *Mol. Gen. Genet.* 134:133-41
17. Cohen, S. N., Chang, A. C. Y., Hsu, L. 1972. *Proc. Natl. Acad. Sci. USA* 69: 2110-14
18. Cohen, S. N., Chang, A. C. Y., Boyer, H. W., Helling, R. B. 1973. *Proc. Natl. Acad. Sci. USA* 70:3240-44
19. Collins, C. H., Hartley, E. G., Pilsworth, R. 1974. *The Prevention of Laboratory Acquired Infection*. Public

532 CURTISS

- Health Lab. Serv., Monogr. Ser. No. 6.
20. Cooke, E. M. 1974. *Escherichia coli and Man*. London: Churchill-Livingston. 96 pp.
 21. Cosloy, S. D., Oishi, M. 1973. *Proc. Natl. Acad. Sci. USA* 70:84-87
 22. Curtiss, R. III. 1965. *J. Bacteriol.* 89: 28-40
 23. Curtiss, R. III. 1969. *Ann. Rev. Microbiol.* 23:69-136
 24. Curtiss, R. III, Caro, L. G., Allison, D. P., Stallions, D. R. 1969. *J. Bacteriol.* 100:1091-1104
 25. Curtiss, R. III, Charamella, L. J., Stallions, D. R., Mays, J. A. 1968. *Bacteriol. Rev.* 32:320-48
 26. Darlow, H. M. 1969. In *Methods in Microbiology*, ed. J. R. Norris, D. W. Robbins, pp. 169-204. New York: Academic. 504 pp.
 27. *Design Criteria for Viral Oncology Research Facilities* 1975. US Dept. HEW, Public Health Serv., Nat. Instit. Health, DHEW Publ. No. (NIH) 75-891
 28. Dimmick, R. L., Vogl, W. F., Chatigny, M. A. 1973. In *Biohazards in Biologic Research*, ed. A. Hellman, M. N. Oxman, R. Pollack, pp. 246-66
 29. Enquist, L., Tiemeier, D., Leder, P., Weisberg, R., Sternberg, N. 1976. *Nature* 259:596-98
 30. Falkinham, J. O. III, Curtiss, R. III. 1976. *J. Bacteriol.* In press
 31. Falkow, S. 1975. *Infectious Multiple Drug Resistance*. London: Pion. 300 pp.
 32. Glover, D. M., White, R. L., Finnegan, D. J., Hogness, D. S. 1975. *Cell* 5:149-55
 33. Griffith, F. 1928. *J. Hyg.* 27:113-59
 34. Hayes, W. 1968. *The Genetics of Bacteria and their Viruses*. New York: Wiley. 925 pp. 2nd ed.
 35. Hedgpeth, J., Goodman, H. M., Boyer, H. W. 1972. *Proc. Natl. Acad. Sci. USA* 69:3448-52
 36. Heffron, F., Sublett, R., Hedges, R. W., Jacob, A., Falkow, S. 1975. *J. Bacteriol.* 122:250-56
 37. Helling, R. B. 1975. *Stadler Symp.* 7:15-36
 38. Hellman, A., Oxman, M. N., Pollack, R., eds. 1973. *Biohazards in Biological Research*. Cold Spring Harbor, NY: Cold Spring Harbor Lab. 369 pp.
 39. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A., Helinski, D. R. 1974. *Proc. Natl. Acad. Sci. USA* 71:3455-59
 40. Hu, S., Ohtsubo, E., Davidson, N., Saeidler, H. 1975. *J. Bacteriol.* 122:764-75
 41. Hu, S., Ptashne, K., Cohen, S. N., Davidson, N. 1975. *J. Bacteriol.* 123: 687-92
 42. Jackson, D. A., Symons, R. H., Berg, P. 1972. *Proc. Natl. Acad. Sci. USA* 69: 2904-9
 43. Jacob, A. E., Hobbs, S. J. 1974. *J. Bacteriol.* 117:360-72
 44. Jacob, F., Wollman, E. L. 1956. *Ann. Inst. Pasteur* 91:486-510
 45. Kedes, L. H., Chang, A. C. Y., Houseman, D., Cohen, S. N. 1975. *Nature* 255:333-38
 46. Krčmery, V., Rosival, L., Watanabe, T., eds. 1972. *Bacterial Plasmids and Antibiotic Resistance*. Prague: Avicenum. 452 pp.
 47. *Laboratory Safety at the Center for Disease Control*, September 1974. US Dept. HEW, Publ. No. (CDC) 75-8118
 48. Lederberg, J. 1975. *Prism* November:33-37
 49. Lennette, E. H., Spaulding, E. H., Truant, J. P., eds. 1974. *Manual of Clinical Microbiology*. Washington DC: Am. Soc. Microbiol. 970 pp. 2nd ed.
 50. Lieb, M. 1970. *J. Virol.* 6:218-25
 51. Lobban, P. E., Kaiser, A. D. 1973. *J. Mol. Biol.* 78:453-71
 52. Mandel, M., Higa, A. 1970. *J. Mol. Biol.* 53:159-62
 53. Matsubara, K., Kaiser, A. D. 1968. *Cold Spring Harbor Symp. Quant. Biol.* 33:769-75
 54. Mertz, J. E., Davis, R. W. 1972. *Proc. Natl. Acad. Sci. USA* 69:3370-74
 55. Mitsuhashi, S. 1971. *Transferable Drug Resistance Factor R*. Tokyo: Univ. Park Press. 203 pp.
 56. Mitsuhashi, S., Hashimoto, H. 1975. *Microbial Drug Resistance*. Tokyo: Univ. Tokyo Press. 566 pp.
 57. Monner, D. A., Jonsson, S., Boman, H. G. 1971. *J. Bacteriol.* 107:420-32
 58. Morrow, J. F., Cohen, S. N., Chang, A. C. Y., Boyer, H. W., Goodman, H. M., Helling, R. B. 1974. *Proc. Natl. Acad. Sci. USA* 71:1743-47
 59. Murray, N. E., Murray, K. 1974. *Nature* 251:476-81
 60. *National Cancer Institute Safety Standards for Research Involving Oncogenic Viruses*, October 1974. US Dept. HEW, Publ. No. (NIH) 75-790
 61. *National Institutes of Health Biohazards Safety Guide* 1974. US Dept. HEW, Public Health Serv. Natl. Inst. Health. GPO Stock No. 1740-00383
 62. Novick, R. P., Morse, S. I. 1967. *J. Exp. Med.* 125:45-59
 63. Novick, R. P., Clowes, R. C., Cohen, S. N., Curtiss, R. III., Datta, N., Fal-

GENETIC MANIPULATION OF MICROORGANISMS 533

- 123: kow, S. 1976. *Bacteriol. Rev.* In press
- rg, P. 64. Parkinson, J. S. Cited by Hershey, A. D., Dove, W. 1971. In *The Bacteriophage λ*, ed. A. D. Hershey, pp. 3-11. Cold Spring Harbor, NY: Cold Spring Harbor Lab. 792 pp.
- 4 69: 65. Ptashne, K., Cohen, S. N. 1975. *J. Bacteriol.* 122:776-81
- Bac- 66. Rambach, A., Tiollais, P. 1974. *Proc. Natl. Acad. Sci. USA* 71:3927-30
- Ann. 67. Reiner, A. M. 1974. *J. Bacteriol.* 119:183-91
- ouse- 68. Schlessinger, D., ed. 1975. *Microbiology—1974*. Washington DC: Am. Soc. Microbiol. 314 pp.
- ature 69. Sgaramella, V. 1972. *Proc. Natl. Acad. Sci. USA* 69:3389-93
- nabe, 70. Shapton, D. A., Board, R. G., eds. 1972. *Safety in Microbiology*. NY: Academic
- s and 71. Shimada, K., Weisberg, R. A., Gottesman, M. E. 1972. *J. Mol. Biol.* 63:483-503
- vice- 72. Signer, E. R. 1969. *Nature* 223:158-60
- r for 73. Singer, M., Söll, D. 1973. *Science* 181:1114
- l. US 74. Sinsheimer, R. 1975. *New Sci.* 16: 148-51
- 8118 75. Skurray, R. A., Hancock, R. E. W., Reeves, P. 1974. *J. Bacteriol.* 119: 726-35
- ovem- 76. Smith, H. W. 1975. *Nature* 255:500-2
- Tru- 77. Stallions, D. R., Curtiss, R. III. 1972. *J. Bacteriol.* 111:294-95
- inical 78. Steere, N. V., ed. 1971. *Handbook of Laboratory Safety*. Cleveland: Chemical Rubber Co. 2nd ed.
- Am. 79. Tanaka, T., Weisblum, B. 1975. *J. Bacteriol.* 121:354-62
73. J. 80. Tanaka, T., Weisblum, B., Schnöss, M., Inman, R. 1975. *Biochem.* 14:2064-72
- Mol. 81. Thomas, M., Cameron, J. R., Davis, R. W. 1974. *Proc. Natl. Acad. Sci. USA* 71:4579-83
1968. 82. Timmis, K., Cabello, F., Cohen, S. N. 1974. *Proc. Natl. Acad. Sci. USA* 71:4556-60
- Biol. 83. Weinsink, P. C., Finnegan, D. J., Donelson, J. E., Hogness, D. S. 1974. *Cell* 3:315-25
- Proc. 181:1114
- Drug 74. Sinsheimer, R. 1975. *New Sci.* 16: 148-51
- Park 75. Skurray, R. A., Hancock, R. E. W., Reeves, P. 1974. *J. Bacteriol.* 119: 726-35
1975. 76. Smith, H. W. 1975. *Nature* 255:500-2
- okyo: 77. Stallions, D. R., Curtiss, R. III. 1972. *J. Bacteriol.* 111:294-95
- oman, 78. Steere, N. V., ed. 1971. *Handbook of Laboratory Safety*. Cleveland: Chemical Rubber Co. 2nd ed.
- 32 79. Tanaka, T., Weisblum, B. 1975. *J. Bacteriol.* 121:354-62
- hang, 80. Tanaka, T., Weisblum, B., Schnöss, M., Inman, R. 1975. *Biochem.* 14:2064-72
- in, H. 81. Thomas, M., Cameron, J. R., Davis, R. W. 1974. *Proc. Natl. Acad. Sci. USA* 71:4579-83
- Natl. 82. Timmis, K., Cabello, F., Cohen, S. N. 1974. *Proc. Natl. Acad. Sci. USA* 71:4556-60
- l. Na- 83. Weinsink, P. C., Finnegan, D. J., Donelson, J. E., Hogness, D. S. 1974. *Cell* 3:315-25
- Stan- 181:1114
- ogenic 74. Sinsheimer, R. 1975. *New Sci.* 16: 148-51
- HEW, 75. Skurray, R. A., Hancock, R. E. W., Reeves, P. 1974. *J. Bacteriol.* 119: 726-35
- azards 76. Smith, H. W. 1975. *Nature* 255:500-2
- HEW. 77. Stallions, D. R., Curtiss, R. III. 1972. *J. Bacteriol.* 111:294-95
- lealth. 78. Steere, N. V., ed. 1971. *Handbook of Laboratory Safety*. Cleveland: Chemical Rubber Co. 2nd ed.
- l. Exp. 79. Tanaka, T., Weisblum, B. 1975. *J. Bacteriol.* 121:354-62
- Cohen, 80. Tanaka, T., Weisblum, B., Schnöss, M., Inman, R. 1975. *Biochem.* 14:2064-72
- ., Fal- 81. Thomas, M., Cameron, J. R., Davis, R. W. 1974. *Proc. Natl. Acad. Sci. USA* 71:4579-83

TAB HH

ERMEK

aving a
(63), in-
EF-Tu-
alogous
, which
leotide
the two
lecules,
national
ing the
leading

of the
ction of
gnition-
arily to
A-site.
osome
sequent
step is
atly, on
nulated
e codon
ognition

).

, PNAS

Synthetic Oligodeoxynucleotides for Analyses of DNA Structure and Function

RAY WU AND
CHANDER P. BAHL

*Section of Biochemistry
Molecular and Cell Biology
Cornell University
Ithaca, New York*

SARAN A. NARANG

*Division of Biological Sciences
National Research Council of
Canada
Ottawa, Canada*

| | |
|--|-----|
| I. Introduction | 102 |
| II. Chemical Synthesis of Oligodeoxyribonucleotides | 103 |
| A. Phosphodiester Method | 103 |
| B. Phosphotriester Method | 105 |
| III. Enzymic Synthesis or Joining of Oligodeoxynucleotides, and the Use of Synthetic Oligonucleotides for Studying Enzymes | 108 |
| A. DNA Ligase | 108 |
| B. DNA Polymerase | 109 |
| C. RNA Ligase | 110 |
| D. Polynucleotide Phosphorylase | 110 |
| E. Other Enzymes Studied with the Aid of Synthetic Oligodeoxynucleotides | 110 |
| IV. Synthetic Oligodeoxyribonucleotides and DNA Sequence Analysis | 111 |
| A. Primer Extension Method for DNA Sequence Analysis | 111 |
| B. Synthetic Oligodeoxyribonucleotides as Tools for Developing New Methods for DNA Sequence Analysis | 114 |
| C. Synthetic Oligodeoxyribonucleotides as Standards for Developing Methods for DNA Sequence Analysis | 116 |
| V. Synthetic Oligodeoxyribonucleotides and the Study of Genes | 119 |
| A. Total Synthesis of Genes | 119 |
| B. Synthetic Oligodeoxyribonucleotides as Probes for Genes | 123 |

| | |
|---|-----|
| VI. Synthetic Oligodeoxyribonucleotides for Studying Protein-DNA Interactions | 125 |
| A. The Lactose Operator-Repressor System | 126 |
| B. The <i>E. coli</i> Tyrosine tRNA Gene-Promoter System | 128 |
| C. The Restriction Endonuclease System | 129 |
| VII. Synthetic Oligodeoxyribonucleotides as Tools in Molecular Cloning of DNA | 130 |
| A. A Specific Method Involving Synthetic Cohesive Ends | 131 |
| B. A General Method Involving Connectors | 133 |
| C. A General Method Involving Adaptors | 134 |
| D. Synthesis of Connectors and Adaptors of Different Lengths | 136 |
| VIII. Concluding Remarks | 136 |
| References | 137 |

I. Introduction

Synthetic oligonucleotides have played a significant role in our understanding of many current problems in molecular biology. Synthetic homopolymeric oligodeoxynucleotides have been extensively used for studying the mechanism of action of various enzymes on DNA synthesis *in vitro* (1). Synthesis of all of the possible 64 triribonucleoside diphosphates have been accomplished for the elucidation of the genetic code (2, 3). More recently, several segments of DNA possessing important biological functions have been chemically synthesized. These include two transfer RNA genes (4, 5), an RNase S-peptide gene (6), an angiotensin hormone gene (7), several lactose operator DNA segments (8-13), and several restriction endonuclease recognition sequences (11, 14-16). The availability of these synthetic molecules has given us the opportunity to understand these systems in greater depth. Short segments of oligodeoxynucleotides have also been synthesized as DNA primers for sequence analysis (17, 18).

In this review we present a general picture of how synthetic oligonucleotides have been used to solve various problems in molecular biology. To keep the length of the article within reasonable limits we have confined ourselves to the area of oligodeoxyribonucleotides of defined sequences, a field closer to our own research interests. The literature surveyed includes articles published before May, 1977. Several review articles covering chemical synthesis (18-21) and certain biological properties of synthetic oligonucleotides (1, 22-27) have appeared.

II. Chemical Synthesis of Oligodeoxyribonucleotides

The synthesis of oligonucleotides involves the joining of nucleotides through phosphodiester linkages. During synthesis, internucleotidic bonds may be maintained either as phosphodiester or phosphotriesters (Fig. 1). Both methods have been used extensively for the synthesis of oligonucleotides with defined sequences.

Synthesis of a dinucleotide containing a (3' → 5') internucleotide linkage was achieved for the first time in 1955 by condensing 3'-O-acetylthymidine with 5'-O-acetylthymidine 3'-benzylphosphochloridate (28). However, various difficulties were encountered in the preparation and purification of the neutral intermediate containing the phosphotriester functions. In 1958, Khorana *et al.* (29) achieved the synthesis of thymidyl-(3' → 5')-thymidine in high yield by introducing a new condensing agent, dicyclohexylcarbodiimide (DCC) for the condensation of 3'-O-acetylthymidine 5'-phosphate with 5'-O-tritylthymidine.

A. Phosphodiester Method

1. BASIC APPROACH

The main feature of this approach is outlined in Fig. 2. DCC was the first condensing agent used. However, for the synthesis of longer chains, mesitylenesulfonyl chloride (30) and triisopropylbenzenesulfonyl chloride (31) were introduced. In the synthesis of an oligonucleotide containing bases other than thymidine, it became important that functional groups such as the free hydroxyls of sugar moieties and the amino groups of adenine, cytosine and guanine be protected. To

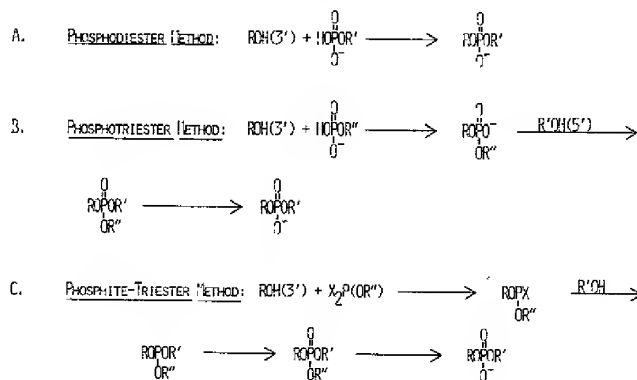


FIG. 1. A scheme showing three methods for the chemical synthesis of oligodeoxynucleotides.

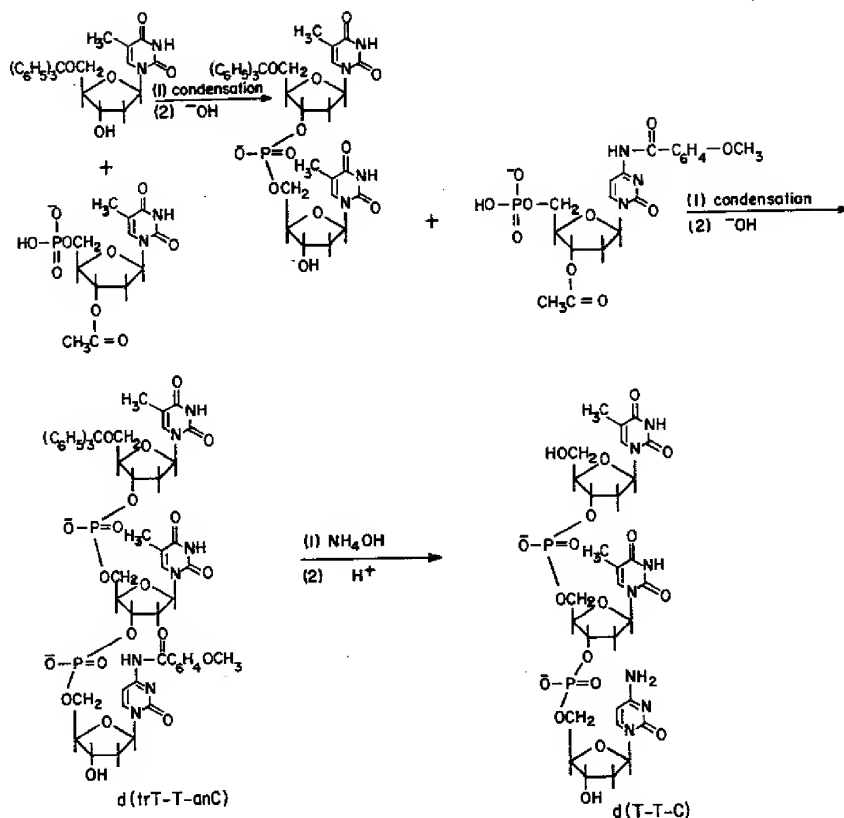


FIG. 2. The phosphodiester method of oligonucleotide synthesis. The synthesis of a trinucleotide is shown.

protect the 5'-hydroxyl group of a deoxynucleoside, *p*-monomethoxy- or dimethoxytrityl groups were used. To protect the amino group of the heterocyclic bases, acylation of the deoxynucleoside followed by selective hydrolysis was employed (32). Using this approach, the synthesis of oligodeoxynucleotides containing homo- or repeating mixed polymers was achieved using chemical polymerization (33), whereas the synthesis of defined sequences of oligonucleotides was accomplished by stepwise block condensation (34). Variations of the phosphodiester approach have been employed for the synthesis of several oligodeoxynucleotides of biological significance. These include sequences corresponding to DNA coding for a derivative of the S-peptide of ribonuclease A (6), the hormone angiotensin II (7), fragments containing natural sequences of the phage T4-gene (35-38), and the λ -endolysin gene (39); and, most notably, all the fragments for the alanine tRNA gene, (4) and for the tyrosine tRNA gene and its control regions (5, 40).

Although the phosphodiester method is useful and reliable, it is inherently laborious and gives low yields of the longer oligodeoxynucleotides for the following reasons. The synthetic intermediates are ionic diesters, insoluble in most organic solvents, and the common method of isolation involves prolonged chromatography on DEAE-cellulose column. A more serious problem is caused by the presence in intermediate fragments of the nucleophilic phosphodiester function, which becomes activated by the condensing reagents (such as arylsulfonylchloride) and causes chain breakage (41). In order to obtain reasonable yields, excess of the incoming nucleotide block must be used, making the diester method inefficient.

Several recent improvements have been made in the diester method to facilitate the isolation of the product. These include aromatic protecting groups to exploit affinity columns (42), solvent extraction (43), "high-pressure" liquid chromatography (44), and reverse-phase "high-pressure" liquid chromatography (45).

2. POLYMER-SUPPORT SYNTHESIS

In spite of the great success of the polymer-support method in peptide synthesis, little related progress has been reported in oligonucleotide synthesis. This is because of the decreased yield at the condensation step with increasing chain length. Recently, the synthesis of a heptanucleotide has been achieved (46) in modest yields by using a cross-linked polydimethylacrylamide resin. This resin contains the amino group of β -alanine residues present as butoxycarbonyl derivatives, which serve as anchoring points. Next, it is treated with β -hydroxythioether, allowing the substituted resin to act as a phosphate protecting group. The chain extension reactions follow a phosphodiester approach by reaction of the support with pyridinium 3'-O-acetylnucleoside 5'-phosphates preactivated with isopropyl benzensulfonyl chloride. This method has undergone further improvements by using a modified Beckman 990 Peptide Synthesizer, resulting in a cycle time of 24 hours per nucleotide addition. Two nonadeoxyribonucleotides, d(pT-T-C-T-G-T-T-G-A) and d(pG-G-A-G-G-A-G-A-A), required as primers for sequencing mouse immunoglobulin mRNA, have been prepared by this approach (Gait and Sheppard, personal communication).

B. Phosphotriester Method

1. EARLIER METHODS

The synthesis of the first dinucleotide was achieved by the triester method by Michelson and Todd (28) in 1955, although the yield was

poor. In 1967, the phosphotriester method was used by Letsinger and Ogilvie (48, 49) for the synthesis of a hexathymidylate in a reasonable yield. This method has been extensively studied in various laboratories (50-54). However, no biologically important oligodeoxynucleotides had been synthesized by this method until 1973, when important modifications were introduced (55). Since then, the modified phosphotriester method has become a more convenient method for synthesis of oligonucleotides of defined sequences.

The triester method involves phosphorylation of the 3'-hydroxy group of a 5'-protected mononucleoside with a phosphate monoesterified by some group such as β -cyanoethyl, *o*-chloro, or *p*-chlorophenyl, followed by subsequent condensation with the primary 5'-hydroxyl group of a suitably 3'-protected nucleoside. Since each internucleotide bond in the product is present as a triester function, isolation and purification by the more conventional techniques of organic solvent extraction and silica-gel chromatography became possible. However, it was observed that, owing to incomplete phosphorylation in the first stage, subsequent coupling with 5'-protected nucleoside led to a complicated reaction mixture. Since these mixtures could not be completely resolved on conventional silica-gel columns, the advantages of large-scale synthesis and high yields of product were somewhat nullified.

2. MODIFIED PHOSPHOTRIESTER METHOD

To overcome the above difficulty, Narang *et al.* modified (55) the "one-pot" triester approach of phosphorylation and coupling (48) to a "two-step" sequential coupling procedure. The basic feature of this procedure is the starting of synthesis of an oligodeoxynucleotide from fully protected mononucleosides containing a fully masked 3'-phosphate group (Fig. 3). Since the resulting intermediate oligonucleotides synthesized contain a fully masked 3'-phosphate group, the necessity for phosphorylation at each condensation stage is eliminated. A similar modification has been employed by Cramer *et al.* (56).

Phosphorylation. The initial phosphorylation of the 3'-hydroxy component was carried out with *bis*(triazoyl)-*p*-chlorophenyl phosphate (57) followed by the addition of β -cyanoethanol.

Coupling Reagents. For the triester method, triisopropylbenzenesulfonyl chloride was the condensing reagent initially used (59). The introduction of arylsulfonyls activated by triazole (8, 58), nitrotriazole (Narang and Stawinsky, unpublished), and tetrazole (60), resulted in higher yields with a minimum of side products. The corresponding imidazolides (61) and nitroimidazolides (62) for the synthesis of oligonucleotides have also been used.

Synthesis of Oligodeoxynucleotides. A 5'-dimethoxytrityl monodeoxynucleoside

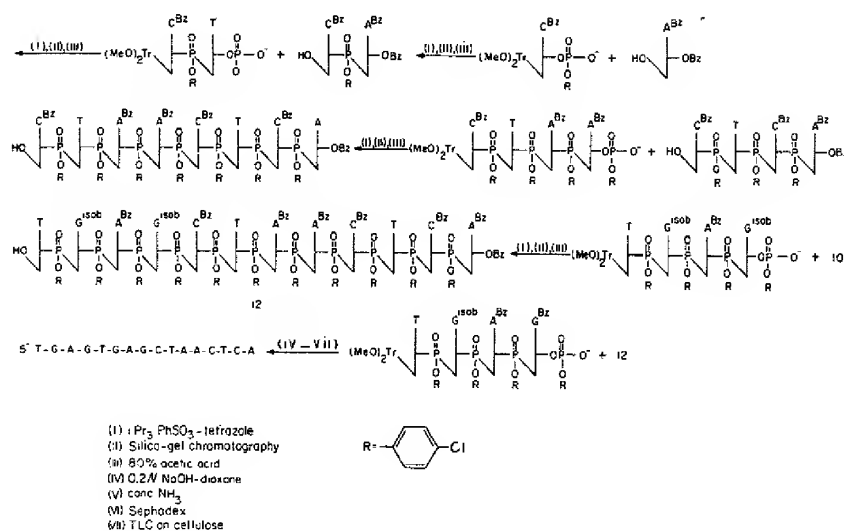


FIG. 3. The modified phosphotriester method of oligodeoxynucleotide synthesis. Triisopropylbenzenesulfonyltetrazole ($\text{iPr}_3\text{PhSO}_3\text{-tetrazole}$) is the condensing agent.

3'-cyanoethyl *p*-chlorophenyl phosphate is the starting material for oligonucleotide synthesis in the triester method. It was essential to develop conditions for the selective removal of the acid-labile group without depurination and for removal of the β -cyanoethyl group without affecting the internucleotide phosphate protecting group, i.e., the *p*-chlorophenyl group. In the first case, we found that benzenesulfonic acid in chloroform (11) selectively removed monomethoxy- and dimethoxytrityl groups at 0°C with a minimum of depurination. Mild treatment with sodium hydroxide/dioxane/water was found to be satisfactory for removal of the β -cyanoethyl groups only in smaller oligonucleotides, while treatment with anhydrous triethylamine/pyridine (63) was found to be highly effective for longer oligonucleotides. A similar observation has been reported independently (64) for the selective removal of β -cyanoethyl groups from fully protected monoribonucleotides. After each condensation step, the reaction mixture was chromatographed on a silica-gel column. After complete deblocking under milder conditions (11), the isolation of the desired product was carried out on a Sephadex gel-filtration column followed by polyethyleneimine-cellulose thin-layer chromatography (65).

3. PHOSPHITE-TRIESTER METHOD

The key feature of this new approach (66) is the coupling of nucleosides by reaction with a phosphodichlorite (a trivalent phosphorus reagent). At the end of the synthesis, phosphite intermediates are oxidized to phosphotriesters (Fig. 1, C). The potential advantages of this approach are: the short time required for the completion of the reaction, the low temperature for the condensation reaction, the relatively high selectivity of phosphochlorite reagents, and favorable yields.

III. Enzymic Synthesis or Joining of Oligodeoxynucleotides, and the Use of Synthetic Oligonucleotides for Studying Enzymes

The chemical synthesis of oligodeoxynucleotides when used in conjunction with enzymic methods permits the preparation of longer segments and sequence variations with minimal effort. Through the use of these oligonucleotides, a great deal has been learned about the mechanism of enzymic action.

A. DNA Ligase

DNA ligase (polynucleotide ligase) obtained from both uninfected and T4-phage-infected *Escherichia coli* catalyzes the covalent joining of two molecules of oligodeoxynucleotide or DNA. The 5'-P from one molecule is joined to the 3'-OH of another to form a phosphodiester bond (67-69). Since it is difficult and time consuming to synthesize oligodeoxynucleotides longer than 15-20 units, a strategy by which short synthetic oligodeoxynucleotides (usually 8-12 bases in length) are joined with the use of DNA ligase has been devised (4, 70-72) (see Fig. 4). This method was used to construct a double-stranded DNA molecule approaching 200 base-pairs in length that included the *E. coli* precursor tyrosine tRNA gene, together with its promoter and termination regions. A minimum of four overlapping nucleotide pairs are required for the joining of oligodeoxynucleotides with DNA ligase.

A second type of DNA ligase activity, which joins even-ended double-stranded DNA molecules, was discovered as a result of the use of T4 DNA ligase for joining synthetic oligodeoxynucleotides of defined lengths and sequences (73, 74). This "blunt-end" joining of DNA molecules works only for T4 DNA ligase, and does not require over-

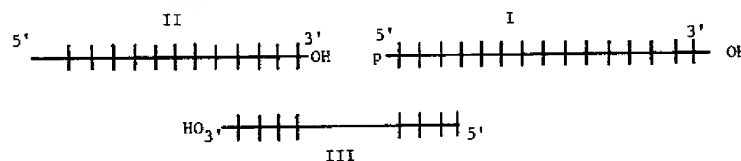


FIG. 4. The joining of short oligodeoxynucleotides with DNA ligase. Each short vertical bar represents a nucleoside. The horizontal lines represent the phosphodiester backbone of the oligodeoxynucleotides. In this scheme, the 3'-OH group of segment II is joined to the 5'-P group of segment I using DNA ligase. Segment III serves as a template for holding segments II and I in adjacent position by base-pairing. Once the joining has been completed, a segment IV can be lined up on the right-hand side of segment III and joined to segment III using segment I as a template. This procedure can be repeated to build a very long sequence.

NARANG

SYNTHETIC OLIGODEOXYNUCLEOTIDES

109

otides,

used in
of longer
ough the
about the

infecting
at joining
from one
hodiester
ynthesize
by which
n length)
)–72) (see
led DNA
ed the *E.*
oter and
tide pairs
with DNA

en-ended
of the use
des of de-
g of DNA
uire over-

OH

. Each short
osphodiester
of segment II
l serves as a
ng. Once the
hand side of
rocedure can

lapping single-stranded regions. Such ligase activity is markedly stimulated by T4 RNA ligase (75) and has been used for joining DNA molecules for molecular cloning (15, 76, 77). (A more extensive discussion on the use of synthetic oligodeoxynucleotides for cloning appears in Section VII.)

B. DNA Polymerase

E. coli DNA polymerase I catalyzes a template-directed "repair" synthesis (78–81) that serves to add specific nucleotides to a primer DNA. At 5°–8°C, the incorporation reaction stops when the single-stranded section is completely repaired (see Fig. 5), and no initiation of new chains occurs. Partial repair synthesis occurs when only one, two, or three of the four deoxynucleoside triphosphates are used.

Repair synthesis can be used to complete a chemically synthesized structure, as shown in Fig. 5, thereby making the synthesis of the 21-nucleotide-long duplex easier. More important, partial repair synthesis on structure I (in Fig. 5) can produce duplex structures of different lengths (e.g., 13, 14, 16, 17, 19, or 21) with relative ease (10). Complete synthesis by chemical means of structures of all intermediary lengths would be extremely time-consuming. In certain molecules, repair synthesis with *E. coli* polymerase failed to incorporate the last nucleotides (71). The use of AMV or RSV reverse transcriptase for repair synthesis resulted in more complete incorporation (10).

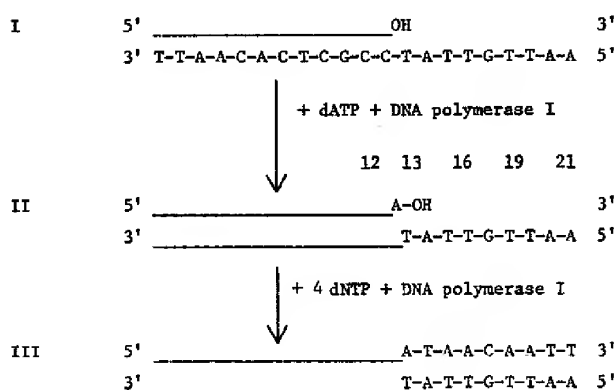


FIG. 5. Repair synthesis of the single-stranded end of a *lac* operator DNA. Structure I represents a 12-nucleotide-long primer hybridized to a 21-nucleotide-long template. When only one deoxynucleoside triphosphate is used for repair synthesis, partial synthesis occurs. This addition of one molecule of pdA gives structure II (a 13-nucleotide duplex). When all four deoxynucleoside triphosphates are used, the repair synthesis is complete and a fully double-stranded DNA molecule is formed (structure III).

Another advantage of repair synthesis is the capability to introduce modified bases at specific locations. For example, if structure II is incubated with 5-Br-dUTP (9) and DNA polymerase I, a 5-Br-dU will be incorporated into position 14 in place of a dT. If 4-S-dTTP (82) is used as the substrate in place of dTTP, structure II (Fig. 5) will contain a 4-S-dT at position 14. The effect of these specific base changes on the biological properties of the DNA can then be tested.

C. RNA Ligase

T4 RNA ligase joins RNA chains in the absence of a template (83). This enzyme also catalyzes the end-to-end joining of single-stranded DNA, a reaction confirmed with the use of synthetic oligodeoxynucleotides (84). The lack of a template requirement makes RNA ligase an attractive reagent for the synthesis of oligodeoxynucleotides with defined sequences. RNA ligase catalyzes the end-to-end ligation of (dT)₁₂₋₁₈ efficiently. However, the rate of this reaction was found to diminish with decreasing chain length of the substrate (84).

D. Polynucleotide Phosphorylase

E. coli polynucleotide phosphorylase catalyzes the addition of a few deoxyribonucleotides to the 3'-OH end of an oligodeoxynucleotide primer (85). Deoxynucleoside 5'-diphosphate, Mn²⁺, and a primer at least three nucleotides in length are required for the reaction (86). This reaction has been used for the stepwise addition of nucleotides to a d(pT-T-A-G) primer to synthesize a tridecanucleotide (87), d(pT-T-A-G-C-A-G-A-A-C-C-G-G), in six steps, as shown in Fig. 6 (1977). In some of the steps, two nucleotides of the same type are added in a single reaction. Because of the large number of steps involved, the overall yield of the final product is rather poor. This method is perhaps more useful in synthesizing variations of a chemically prepared sequence: for example, a specific nonanucleotide can be converted to four different decanucleotides with A, G, T, or C at the 3' end.

E. Other Enzymes Studied with the Aid of Synthetic Oligodeoxynucleotides

A synthetic octanucleotide, d(A-C-C-A-T-C-C-A), was employed to develop a useful method for labeling the 3' ends of oligodeoxynucleotides or DNA. Terminal transferase was used (88, 89) to catalyze the addition of a single [³²P]ribonucleotide to the 3' end of the octanucleotide to give d(A-C-C-A-T-C-C-A)³²p-rN, where N is any one of the four nucleosides. The 3'-labeled oligonucleotide or DNA can then be

RANG

SYNTHETIC OLIGODEOXYNUCLEOTIDES

111

duce
II is
I will
92) is
ntain
on the

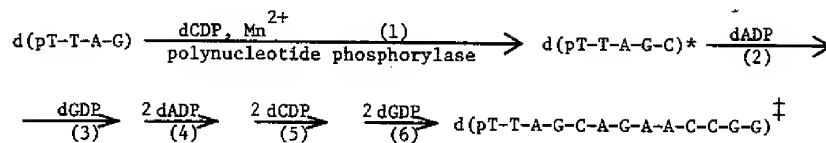


FIG. 6. Enzymic synthesis of oligodeoxynucleotides of defined sequence. * This pentanucleotide was purified on a DEAE-cellulose column to remove the diaddition product, d(pT-T-A-G-C-C), and the starting material. The same column purification was used after each step: (2), (3), (4), (5), and (6). In order to improve the yield, the unreacted primer was isolated at each step and recycled in one or several repeated synthetic steps. † This tridecanucleotide constitutes a segment of the yeast iso-cytochrome *c* gene.

(83).
inded
xynu-
ase an
th de-
on of
nd to

used directly for sequence analysis. Alternatively, it may be used as a primer for further extension with deoxynucleotides. Subsequently, the extended segment may be analyzed after its cleavage from the primer at the ribonucleotide linkage (90-92).

The mechanism of action of *E. coli* exonuclease III was studied by hybridizing (dA)₁₀₀₀ with oligothymidylate labeled at the 3' end with one, two, or more ribonucleotides. It was found that exonuclease III can cleave one or two consecutive ribonucleotides regardless of whether the ribonucleotides are base-paired or mismatched (93).

Oligodeoxynucleotides of defined sequences such as d(³²pA-C-A-A-T-T) and d(³²pA-A-T-T-G-T-G-A-G-C-G-G) have been used to study the 3' → 5' exonuclease activity of *E. coli* DNA polymerase I. It was found that at pH 7.1 and 5°C, DNA polymerase I does not digest d(pA-C-A-A-T-T) but degrades longer oligodeoxynucleotides as well as hexathymidylate to tetranucleotides. This indicates that digestion is dependent on both the size and sequence of the oligonucleotides. At 15°C or higher temperature, every oligonucleotide tested (hexanucleotides or longer oligonucleotides) produced a terminal 5'-dinucleotide as the major product (Bahl and Wu, unpublished).

IV. Synthetic Oligodeoxyribonucleotides and DNA Sequence Analysis

A. Primer Extension Method for DNA Sequence Analysis

The sequence of a DNA molecule from a specific location can be determined by either the end-labeling approach or the primer-extension approach.

1. THE END-LABELING APPROACH

For sequence analysis from the 5' end of a DNA molecule, the 5'-OH group can be labeled with ³²P using [γ-³²P]ATP and

n of a
oxynu-
and a
action
of nu-
eotide
Fig. 6
pe are
ps in-
. This
chemi-
de can
at the

oyed to
oxynu-
yze the
octanu-
e of the
hen be

polynucleotide kinase (94, 95). After labeling, the two 5' ends of a double-stranded DNA can be separated either by strand separation or by cutting the DNA molecule with a restriction enzyme, followed by separation of the two ends (17, 25). Each purified 5'-end-labeled DNA can be sequenced using one of three methods: exhaustive digestion with pancreatic DNase I plus venom phosphodiesterase, or *E. coli* exonuclease I, to determine the sequence of mono- or dinucleotides (17, 25, 94, 96, 97); partial enzymic digestion followed by mobility analysis to determine sequences up to 15 nucleotides in length (39, 98–100); or partial chemical digestion followed by gel electrophoresis to determine up to 120 nucleotides from the labeled 5' end (101).

For sequence analysis from the 3' end of a DNA molecule, the 3' end can be labeled by introducing a single deoxynucleotide (102–105) or a single ribonucleotide (89, 106–108). Sequence analysis can then proceed from the labeled 3' end using the same principles described for analysis from the 5' end.

2. THE PRIMER-EXTENSION APPROACH

For sequence analysis starting from an internal portion of a long DNA molecule, specific primers can be used. The principle for the primer-extension method described by Wu *et al.* (81, 109) is illustrated in Fig. 7. To analyze the DNA sequence from a specific location, such as nucleotide 120, an oligodeoxynucleotide primer is first hybridized onto the template strand starting at nucleotide 120, and then enzymically extended toward nucleotide 240 by repair synthesis (102, 110–113). By determining the sequence of the extended region (dotted line), the sequence of the complementary template strand can be deduced.

The oligodeoxynucleotide primer of defined sequence can be synthesized chemically according to DNA sequences derived from several sources: a known protein sequence coded by the specific region of a gene (39, 109, 110, 114), a known RNA sequence (72, 111, 115–117), or a short known DNA sequence (90, 118, 119).

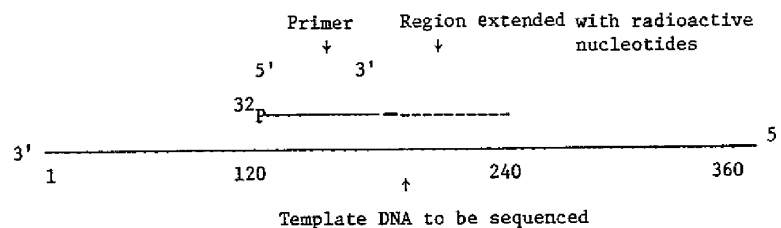


FIG. 7. Primer-extension approach for DNA-sequence analysis.

nds of a ration or owed by ed DNA ligation r *E. coli* leotides mobility ngth (39, ophoresis (101). le, the 3' 102-105) can then escribed

FIG. 8. The plan for sequencing the promoter region of the tyrosine tRNA gene. The l-strand of $\phi 80$ psu^{+}_{III} DNA was used as the template strand for primer elongation. A synthetic 21-nucleotide-long primer was used to prime DNA synthesis by DNA polymerase. With this primer, a sequence up to nucleotide 39 in the promoter region was determined. Then a second synthetic primer in the newly sequenced region was used to determine the rest of the sequence up to nucleotide 59 in the promoter region (143).

A synthetic octanucleotide, d(A-C-C-A-T-C-C-A), was used as a primer to develop a primer-extension method for DNA sequence analysis (110). This method involves the enzymic incorporation of a ribonucleotide at specific locations (ribo-substitution method) to facili-

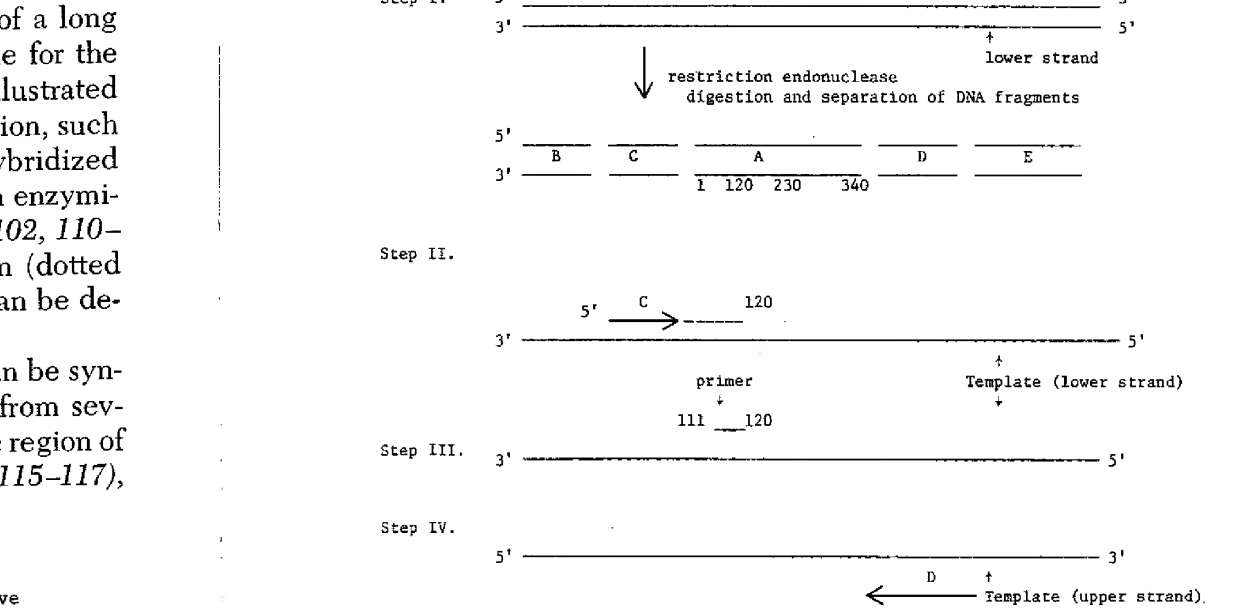


FIG. 9. The use of DNA fragments as primers for sequence analysis. In step I, a double-stranded DNA molecule is digested by a specific restriction endonuclease to produce five specific DNA fragments. The fragments are separated by gel electrophoresis according to size. In step II, the two strands of a double-stranded DNA are separated. For sequence analysis of segment A, the lower strand is used as the template and is hybridized with heat-denatured fragment C. Enzymic extension of fragment C by repair synthesis will allow sequence analysis of the initial part of segment A (e.g., nucleotides 1-120). Steps III and IV, see Section IV, A, 2 in text.

tate the analysis of long sequences. Using fd DNA as template and the synthetic octanucleotide as primer, a sequence of 50 nucleotides was determined.

Synthetic oligodeoxynucleotide primers corresponding to the *E. coli* tyrosine tRNA sequence were employed for sequence analysis. Using the separated strands of $\phi 80$ pSu_{III} DNA as template and synthetic primers, the sequence at the promoter region (90) of the tyrosine tRNA gene was determined as shown in Fig. 8.

Specific DNA fragments generated by the restriction endonuclease digestion of a DNA have proved to be useful primers for DNA sequence analysis (120–122). Sanger *et al.* thus achieved the sequencing of ϕX 174 DNA, a monumental task involving approximately 5000 nucleotides (122–124).

The method described by Sanger and Coulson (113) can be used to determine up to 120 nucleotides starting from a restriction enzyme recognition site (such as the C/A junction shown in Fig. 9). If fragment A is longer than 300 nucleotides and has no other known restriction enzyme recognition sites, then the center region of fragment A can be analyzed by synthesizing a decanucleotide primer corresponding to the known DNA sequence (e.g., nucleotides 111–120 in Fig. 9, step III). Using this synthetic primer, the sequence of fragment A from nucleotide 120 to approximately 230 can be determined. Nucleotides 230–340 can be determined from the opposite direction by using the upper strand as the template and fragment D as a primer (Fig. 9, step IV).

The DNA molecule used for primer-binding and extension may be a single-strand DNA isolated from a double-stranded DNA (101, 115, 120, 125), a double-stranded DNA made partially single-stranded by *E. coli* exonuclease III digestion (38, 39), or a single-stranded circular DNA (110, 121, 123).

B. Synthetic Oligodeoxynucleotides as Tools for Developing New Methods for DNA Sequence Analysis

A synthetic decanucleotide, d(A-G-A-A-A-T-A-A-A-A), was utilized as a primer for developing a rapid new method for DNA sequence analysis, commonly known as the “plus-minus” method (113) (Fig. 10). DNA polymerase I is first used to extend the primer oligonucleotide and copy the template in the presence of the four deoxynucleoside triphosphates, one of which is labeled with ³²P. This synthesis produces oligonucleotides of different lengths, all starting from the 3'-OH end of the primer. The mixture is then passed through a small agarose column to remove the deoxynucleoside triphosphates. The extended primers are subsequently treated in one of two ways.

NARANG

SYNTHETIC OLIGODEOXYNUCLEOTIDES

115

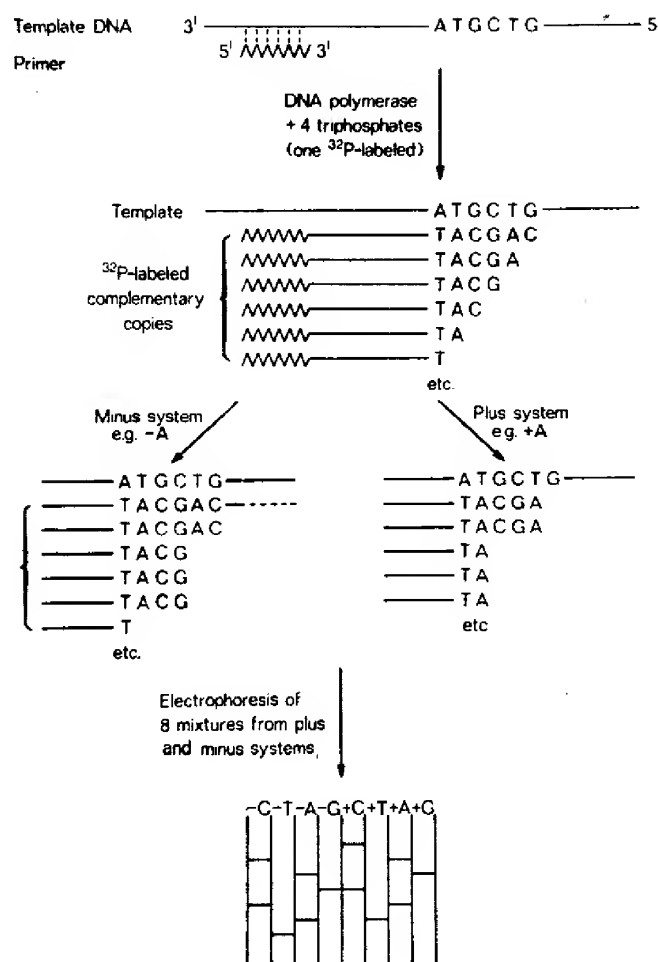


FIG. 10. The principle of the "plus-minus" method of DNA sequence analysis (113).

1. THE MINUS SYSTEM

This system utilizes the partial-repair principle introduced for determining the sequence of "sticky" ends of bacteriophage λ (79). The random mixture of oligonucleotides, still hybridized to the template DNA, is incubated with DNA polymerase I in the presence of three deoxyribonucleoside triphosphates. Synthesis then proceeds as far as it can on each chain and stops when points are reached where the missing triphosphate is required. Thus, if dATP is the missing triphosphate, each chain will terminate at its 3' end at a position preceding a dA residue. Four separate samples are incubated, each with a different one of the four triphosphates missing.

DNA molecules in the four incubation mixtures are then denatured

to separate the newly synthesized strands from the template, and subjected to electrophoresis on polyacrylamide gels containing 8 M urea. This fractionation system arranges all the oligonucleotides having a common 5' end according to chain length. Ideally, each oligonucleotide should be separated from that neighbor differing in length by one nucleotide. The radioautograph from the minus-dA mixture contains bands corresponding to positions preceding the dA residue in the synthesized chain. Reactions involving the other three minus systems are examined in a similar fashion. By comparing the positions of the bands on the radioautographs of the four mixtures, the complete DNA sequence may be deduced. Since this system alone is not sufficient to establish the sequence unambiguously, a second system is used in conjunction with it.

2. THE PLUS SYSTEM

The "plus" system makes use of the property of T4 induced DNA polymerase to degrade double-stranded DNA from its 3' end. By the addition of a single deoxynucleoside triphosphate, the degradation is effectively stopped. This is due to the constant replacement of a nucleotide in the DNA by incorporation of the added nucleotide (103, 126). This method is applied to the random oligonucleotide mixture obtained above. Samples are incubated with T4 DNA polymerase and one of the four triphosphates and then fractionated by electrophoresis on polyacrylamide gel. For example, in a plus-dA mixture, all chains will terminate with a dA. The position of the chains will be indicated by bands on the radioautograph. Usually these bands will be one nucleotide longer than the corresponding bands in the minus-dA system. If there is more than one consecutive dA, the distance between the bands in the minus-dA and plus-dA system will indicate the number of such nucleotides. Thus, by a combined use of the plus and minus systems, it is possible to deduce the sequence of a molecule more than one hundred nucleotides in length starting from a synthetic primer. Using this method together with fragments generated by restriction enzymes, a large portion of the genome of ϕ X174 has been determined (122-124).

C. Synthetic Oligodeoxyribonucleotides as Standards for Developing Methods for DNA Sequence Analysis

The 2-D (two-dimensional) electrophoresis-homochromatography method (127-129) has been extensively used for the fractionation of T1 and pancreatic ribonuclease digestion products of RNA molecules. This method has also been used for the sequence analysis of pyrimidine tracts in DNA up to 20 nucleotides long (98), and sub-

and sub-
3 M urea.
having a
oligonu-
length by
ture con-
lue in the
s systems
ons of the
ete DNA
fficient to
s used in

sequently has been adapted for sequence analysis involving all four nucleotides. However, owing to the overlap between mobility shifts exhibited by different nucleotides (Fig. 11), and the subjectivity of visual inspection, it is often difficult to assign the correct sequence. To solve this problem, a quantitative mobility-shift method using synthetic oligonucleotides of defined sequences as standards has been developed (100). Although the recently developed chemical method of Maxam and Gilbert (101) and the plus-minus method of Sanger and Coulson (113) can determine a long DNA sequence in a relatively short time, the 2-D electrophoresis-homochromatography system is by no means obsolete. This system is especially useful for sequencing the terminal five to ten nucleotides of long DNA fragments that are often difficult to sequence by the chemical method (101), and for determining short sequences at restriction endonuclease recognition sites (108, 130).

1. THE QUALITATIVE 2-D FRACTIONATION SYSTEM

For the fractionation of oligoribonucleotides by the 2-D method (127-129), the first dimension is electrophoresis on a cellulose acetate strip at pH 3.5, which separates nucleotides on the basis of charge. The second dimension is homochromatography on DEAE-cellulose thin-layer plates, which separates nucleotides on the basis of length. As shown in Fig. 11, the addition of pdC in this system (98) causes a shift to the right. The addition of a pdA causes no or only a slight shift

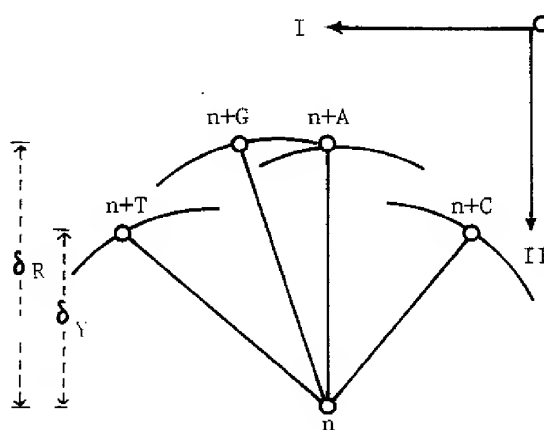


FIG. 11. The usual pattern of mobility-shift of an oligodeoxyribonucleotide (n) extended by a single nucleotide on two-dimensional (2-D) electrophoresis-homochromatography. The first dimension is electrophoresis on cellulose acetate at pH 3.5, and the second dimension is homochromatography; δ_R and δ_Y represent the distances for addition of purine nucleotides and pyrimidine nucleotides, respectively, fractionated in the second dimension.

to the left or right, while the addition of pdT or pdG causes a larger shift to the left.

This method has been used extensively for sequence analyses of oligodeoxynucleotides. However, in many cases the visual inspection of shifts did not give conclusive results. Unambiguous sequences were obtained only after confirmation by other methods, such as end-group, nearest-neighbor, or pyrimidine-tract analysis. To convert it into an independent quantitative tool, the mobility-shift method was improved by the use of synthetic oligonucleotides of defined sequence to establish quantitative parameters (100). The major improvement was in the way the electrophoretic mobility shifts are interpreted. Instead of relying on visual inspection, the idea of calculating expected mobilities was developed. By comparing the observed with the calculated mobility-shifts for various oligonucleotides, a sequence could be deduced without ambiguity. We have also modified the homomixtures (131) to make more clear the distinction between purine and pyrimidine shifts in the second dimension of this 2-D method.

2. THE QUANTITATIVE 2-D SYSTEM FOR SEQUENCE ANALYSIS (100)

The calculation of electrophoretic mobilities was based on the following considerations. A molecule in a fluid subjected to a voltage gradient E moves with a velocity $U = EQ/K$, where Q is the net charge on the molecule at a given pH, and K is a constant that increases proportionally to the size of the molecule. All mobilities were expressed with respect to an internal standard of [^{14}C]pdT, which has a charge of -1 at pH 2–5. The electrophoretic mobility of an oligodeoxynucleotide on cellogel at pH 3.5 relative to pdT can be written as $U_t = Q/K$. The average K values (K) for every length of nucleotides, n , was determined experimentally with the use of a number of different synthetic oligodeoxynucleotides of defined sequence. Each 5'- ^{32}P -labeled oligomer was partially digested with venom phosphodiesterase. The products of digestion were then fractionated on the 2-D electrophoresis-homochromatography system under identical conditions. As the base composition and sequence of each partial product was known, the net charge $Q = \sum_n q$ could be determined (where q is the charge of each mononucleotide). The observed mobility was calculated by the following relation (100, 132):

$$U_t^{\text{obs}} = \frac{\text{distance traveled by oligomer in the first dimension}}{\text{distance traveled by pdT in the first dimension}}$$

Since both U_t and Q could be measured for any given oligonucleotide, the value of K_n ($n = 1, 2, 3, \dots$) could be obtained through the relation $K_n = Q/U_t$. By determining the K_n values for each oligonu-

JARANG

larger

yses of
ection
es were
-group,
into an
as im-
ence to
ent was
Instead
pected
e calcu-
ould be
homo-
ine and

s (100)

the fol-
voltage
charge
increases
ere ex-
ch has a
of an
can be
ngth of
se of a
ned se-
ed with
ien frac-
system
ence of
ould be
le). The
00, 132):

on
t

oligonu-
through
oligonu-

SYNTHETIC OLIGODEOXYNUCLEOTIDES

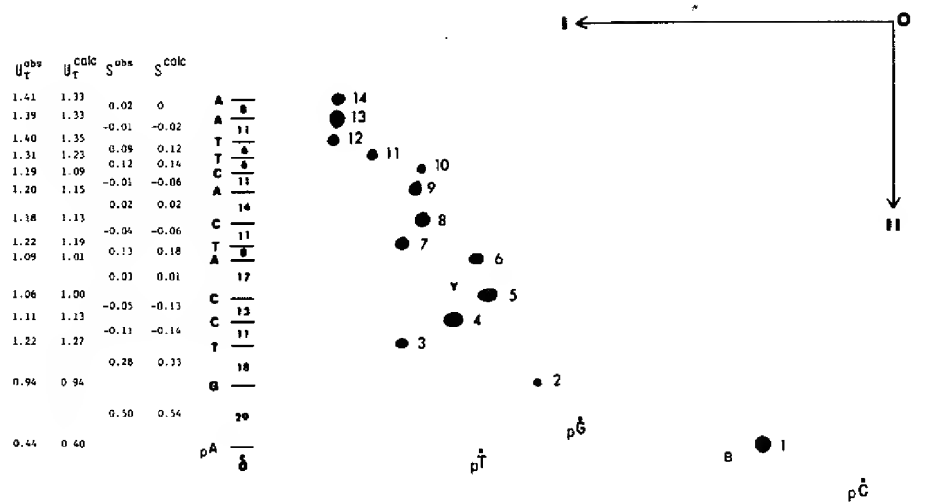


FIG. 12. Two-dimensional electrophoresis-homochromatography for sequence analysis. The 14-mer, d(32pA-G-T-C-C-A-T-C-A-C-T-T-A-A), was partially digested by snake venom phosphodiesterase and electrophoresed on cellulose acetate at pH 3.5. The second dimension is on DEAE-cellulose thin-layer plate with homo-mixture IV (131). The mobility shift is calculated according to Tu *et al.* (100). pT, etc., mark the location of 14C-labeled nucleotides used as markers.

cleotide up to a length of 15 nucleotides, the average K_n value (\bar{K}_n) for each n was obtained (100). Thus, using these average values for K_n , the mobility of any oligonucleotide can be calculated. From the values, the predicted mobility-shift between two adjacent nucleotides n and $n + 1$ differing by one mononucleotide unit can be calculated through the relation $S^{calc} = U_{T,(n+1)} - U_{T,n}^{calc}$. The mobility-shift is characteristic of the mononucleotide unit added to an oligomer of length n to give oligomer of length $n + 1$. By comparing the observed mobility-shift value, $S^{obs} = U_{T,(n+1)}^{obs} - U_{T,n}^{obs}$, with the S^{calc} for the addition of any one of the four mononucleotides, the sequence of the oligomer can be unambiguously derived. An example of the mobility-shift calculation for sequence analysis is shown in Fig. 12. Quantitative mobility-shift analysis has been used extensively in the authors' laboratory for the sequence analysis of tumor virus DNA (108, 132a).

V. Synthetic Oligodeoxyribonucleotides and the Study of Genes

A. Total Synthesis of Genes

The most extensive work in the field of synthetic oligonucleotides has been done in Khorana's laboratory. One of his efforts has been

directed toward the synthesis of biologically functional DNA, i.e., genes capable of giving rise to gene products. This work has resulted in the development of a large number of techniques in chemical and enzymic synthesis of DNA. During the past decade, Khorana *et al.* have reported the synthesis of two transfer RNA genes, one for the yeast alanine tRNA and the other for the *E. coli* tyrosine tRNA (4, 5).

Work on the synthesis of the yeast alanine tRNA gene (4) was initiated in 1965. The choice of this gene for synthesis was based on the fact that it was the first tRNA whose sequence was known (133). Thus, the sequence of the deoxynucleotide in the gene could be derived directly. Moreover, tRNA molecules are interesting in that they are recognized by a large number of components of the protein-synthesizing machinery. These include aminoacyl-tRNA synthetases, the nucleotidyl transferases that repair the C-C-A end of the tRNAs, the ribosomes, and RNAs. The synthesis and study of the tRNA gene would serve to provide useful knowledge about its structure and function. The strategy involved in the synthesis of the 77-base-pair-long duplex DNA structure (Fig. 13) involved three steps. (a) Chemical synthesis of 15 oligodeoxynucleotide segments ranging from 8 to 20 nucleotides in length. These segments were to comprise both strands of the intended DNA. Those belonging to the complementary strands

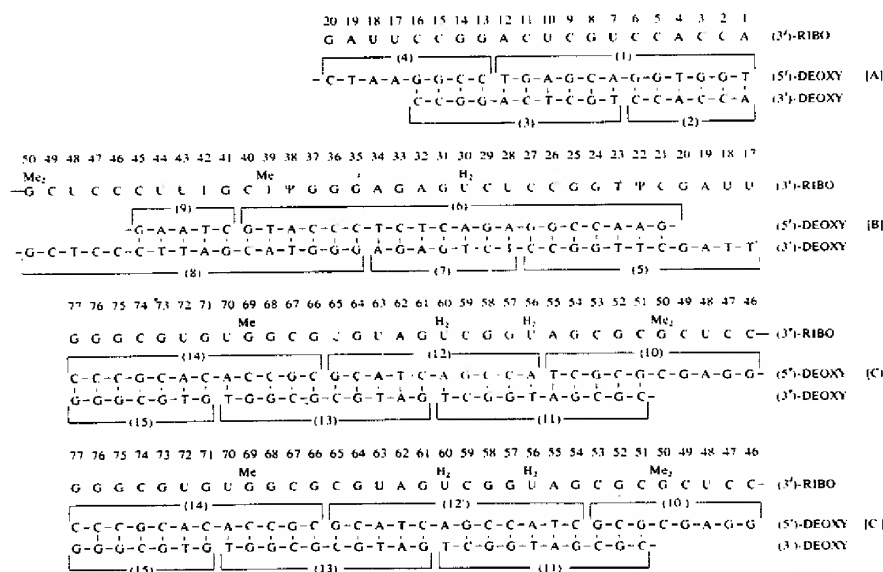


FIG. 13. Plan for the total synthesis of a yeast alanine tRNA gene. The chemically synthesized segments are in brackets, the serial number of each segment being shown within the brackets. A total of seventeen segments (including 10' and 12') varying in chain length from penta to icosanucleotides were synthesized.

SYNTHETIC OLIGODEOXYNUCLEOTIDES

121

had an overlap of four to five nucleotides. (b) The phosphorylation of the 5'-hydroxyl group with [^{32}P]ATP, using T4 polynucleotide kinase. (c) The head-to-tail joining of the properly aligned segments to form bihelical complexes, using the T4 polynucleotide ligase.

After five years of intensive work, the total synthesis of yeast alanine tRNA gene (4) was achieved. Although the synthesis was a great accomplishment and a tribute to the chemical methodology and enzymology developed through this process, attempts to do any biological work with the artificial gene were not successful, owing to its lack of sufficient information for the desired *in vitro* transcription to make alanine tRNA, the gene product. Instead, the *E. coli* DNA-dependent RNA polymerase made only limited amounts of the transcript, which initiated and terminated nonspecifically.

The second gene synthesized by Khorana *et al.* was the *E. coli* tyrosine tRNA gene. The major consideration favoring synthesis of this gene was the relative abundance of information available for the *E. coli* system. This included an understanding of the cell-free protein-synthesizing system, the biochemistry of the ribosomes,¹ and the various factors required for initiation, elongation, and termination of polypeptide chains.² In addition, the aminoacyl-tRNA synthetase corresponding to the tyrosine tRNA (134) had been purified and characterized, and the primary sequence of this tRNA determined by two independent methods (135–138). Extensive genetic and biochemical analysis of this gene has been carried out (139–141). Using the lysogenic bacteriophage $\phi 80$, which contains the *E. coli* tyrosine tRNA suppressor gene, and using suppression as the assay, a large number of mutants involving the tyrosine tRNA gene were isolated. The nature of these mutations and changes in function were studied. The tyrosine tRNA exists *in vivo* as a 126-nucleotide-long precursor containing a 5' triphosphate. This precursor is then processed to give the mature tRNA (142).³

The plan for the chemical synthesis of tyrosine tRNA gene included the sequences shown in Fig. 14. A total of 26 fragments, 4–12 nucleotides long, were synthesized. These fragments were joined together by T4 DNA ligase in a stepwise fashion to get the 126 base-pair-long duplex DNA coding for the precursor tRNA. Single-stranded tails at the ends of the synthetic DNA were included to allow for

¹ See articles by Brimacombe *et al.* in Vol. 18, and by Bermek and by Spirin in this volume (Ed.)

² See article by Grunberg-Manago and Gros in Vol. 20 of this series (Ed.)

³ See article by J. D. Smith in Vol. 16 of this series.

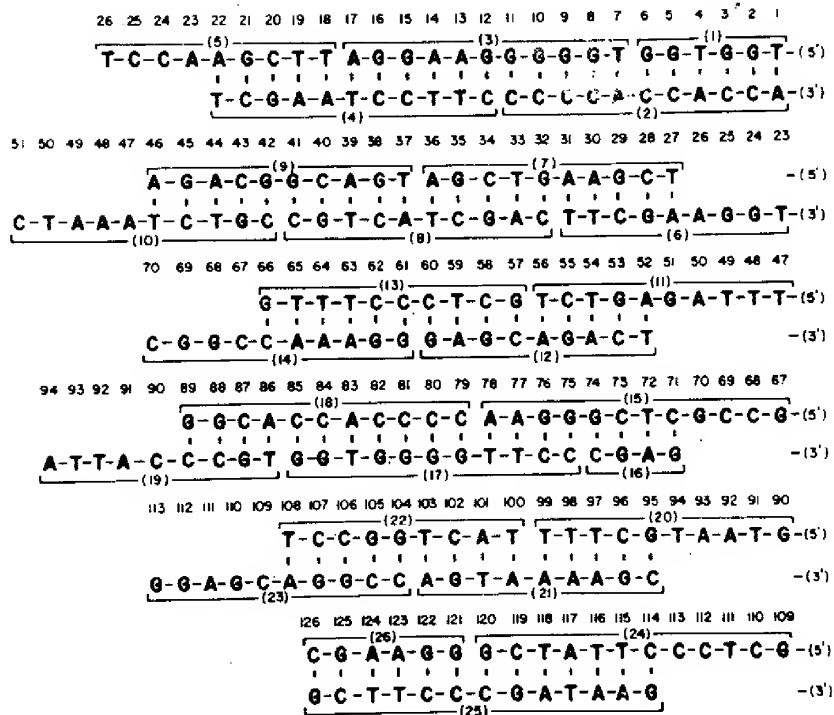


FIG. 14. Plan showing the 26 segments synthesized chemically in the total synthesis of the bihelical DNA corresponding to the precursor for the *Escherichia coli* tyrosine suppressor tRNA gene.

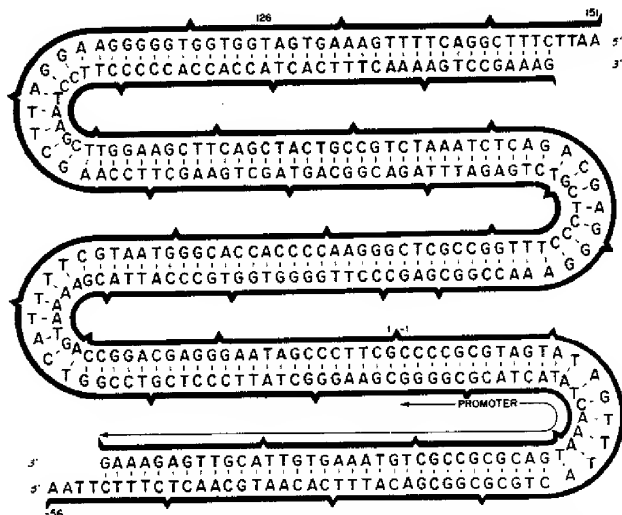


FIG. 15. The sequence of the *Escherichia coli* tyrosine suppressor tRNA gene containing the promoter and the terminator region, synthesized chemically.

extension to the regions comprising the promoter and the terminator region of the genes.

To make the synthetic gene a functional unit, the sequence of the promoter region (56 base-pairs) and the processing terminator region (25 base-pairs) was determined (143) by the primer-extension method, using the $\phi 80$ pSu_{III}⁺ DNA as a template. Finally, the fragments corresponding to these two control regions were synthesized and joined to the structural gene. Thus, the total synthesis of the gene, complete with its promoter and terminator regions, was achieved (Fig. 15) (40).

To test for the biological activity of this gene, *EcoRI* endonuclease recognition sequences were added to both ends of the gene for cloning into *E. coli*. Biological activity was assayed by the ability of the cloned gene to suppress amber mutations *in vivo*. Integration of the synthetic gene into the λ phage genome gave it the ability to form plaques on Su⁻ host bacteria. The presence of the *E. coli* tyrosine suppressor tRNA gene in the recombinant phage was confirmed by its reisolation from phage DNA after amplification. The synthetic gene was also incorporated into the colicin E₁ ampicillin-resistance plasmid. The recombinant plasmid transformed bacteria for both Amp^r and suppressor-plus (su⁺) markers with approximately equal efficiency (144).

In addition to the synthesis of tRNA genes, the DNA coding for biologically active peptides have been synthesized in various laboratories, such as the synthesis of oligodeoxynucleotides corresponding to the bovine insulin chain A (145), to the S-peptide of bovine pancreatic ribonuclease A (6), and to the human peptide hormone angiotensin II (7). Attempts to translate these synthetic DNA molecules into peptides are in progress. If successful, they may help in producing useful peptide products via DNA synthesis.

B. Synthetic Oligodeoxyribonucleotides as Probes for Genes

Synthetic oligonucleotides can also be used as specific probes for studying various aspects of gene structure and function. Synthetic oligonucleotides corresponding to portions of several genes have been synthesized, e.g., λ endolysin (39), T4 lysozyme (35-38), and yeast iso-1-cytochrome *c* gene (87, 117). These oligonucleotides can be used as primers for determining the primary DNA sequence by the primer extension method (see Section IV, A). They can also be used for isolating specific mRNA and for locating the respective genes by the hybridization method of Southern (146).

Detailed study of the yeast iso-1-cytochrome *c* gene, using a synthetic oligonucleotide as a probe, is in progress in the author's labora-

tory. A pentadecadeoxynucleotide (15-mer) primer, 5'd(A-G-C-A-C-C-T-T-T-C-T-T-A-G-C), corresponding to amino acids 8 to 12 of yeast iso-1-cytochrome *c*, has been synthesized. Using the mRNA as a template and extending the primer by DNA-polymerase-catalyzed repair synthesis, a sequence of 20 nucleotides at the 5' end of the AUG initiation codon was obtained (Szostak, Stiles and Wu, unpublished).

The 15-mer has been chemically linked to cellulose, an immobilized support, with the condensing agent dicyclohexylcarbodiimide, and used to purify and isolate the iso-1-cytochrome *c* mRNA (Szostak *et al.*, unpublished). A hexanucleotide, d(A-C-C-A-G-C), which corresponds to the two major codons of *Bombyx mori* fibroin mRNA, GCU (alanine) and GGU (glycine), has been synthesized (147). An affinity column made of polymerized hexanucleotide will preferentially trap the mRNA. The availability of pure mRNA should facilitate work in this system.

A labeled 15-mer corresponding to the cytochrome *c* gene has also been utilized for identifying the restriction fragments containing the iso-1-cytochrome *c* gene. Yeast DNA was digested with either *Eco*RI or *Bam*I restriction enzyme and fractionated on an agarose gel. After transferring the DNA from the gel to a nitrocellulose filter (146), ³²P-labeled 15-mer or an enzymically extended 15-mer (cDNA) was used as a probe to hybridize with the cytochrome *c* gene. A band presumably containing the iso-1-cytochrome *c* gene was detected on the gel. With this synthetic probe, it was found that in yeast with a wild-type iso-1-cytochrome gene the *Eco*RI DNA fragment carrying the gene has an M_r of 4.5×10^6 . A mutant with a mutated *Eco*RI site in the cytochrome *c* gene the size of the *Eco*RI fragment is about 6×10^6 (Szostak, Stiles, Sherman and Wu, unpublished).

The labeled 15-mer has also proved useful in screening plasmids for the cytochrome *c* gene from a colony bank of yeast DNA. Once the yeast iso-1-cytochrome *c* gene is cloned, large amounts of the cytochrome *c* DNA can be isolated. For sequence analysis, the synthetic 15-mer functions as a primer to be annealed with the cloned cytochrome *c* DNA (strand separation may or may not be necessary) and extended with the use of *E. coli* DNA polymerase I. Then either the method of Sanger and Coulson (113) or that of Maxam and Gilbert (101) can be used for sequence analysis.

Using the same principle as in the probing and analysis of the yeast cytochrome *c* gene (117), synthetic oligonucleotides with defined sequences can be very valuable as hybridization probes for the isolation and cloning of single-copy genes in mammalian cells. For single-copy

ANG
-C-
ast
s a
zed
of
Vu,
im-
car-
e c
-A-
lori
the-
ide
NA

also
the
oRI
fter
32p-
sed
um-
gel.
ype
has
yto-
tak,

nids
the
yto-
netic
yto-
and
the
bert

east
l se-
tion
copy

genes that produce appreciable amounts of mRNA (e.g., more than 10% of the total cellular mRNA), the specific mRNA can be purified, copied into cDNA and double-stranded DNA, and then cloned (148, 149). However, for genes that produce small amounts of mRNA (less than 1% or 0.1% mRNA of the total mRNA) the above method is not usable. There are two approaches to the cloning of these single-copy genes. An oligonucleotide approximately 18–20 nucleotides in length can be synthesized and used as a primer and hybridized to the intended specific mRNA in a mixture of mRNAs from a mammalian cell. Following enzymic extension of the synthetic primer, the resulting cDNA can be made double-stranded and cloned (148, 149). Alternatively, a given DNA may first be digested by a restriction enzyme and fractionated on an agarose gel. The synthetic oligonucleotide may then be used as a probe for hybridization with a specific gene in the original DNA. Next, the DNA can be cloned, and the transformed cells may be screened for the desired gene by hybridization with the same synthetic probe (150).

VI. Synthetic Oligodeoxyribonucleotides for Studying Protein–DNA Interactions

One of the most basic problems in molecular biology is understanding protein–nucleic acid interactions; i.e., how do the specific amino acids of proteins interact with the nucleotide sequences of the DNA? The sequences of a number of interesting segments of DNA such as operator, promoter, and restriction endonuclease recognition sites have been elucidated in recent years. For studying the point-to-point interaction of nucleic acids with proteins, large amounts of the DNA fragments are required. DNA fragments may be isolated following suitable restriction enzyme digestion, incorporation into bacterial plasmids, and amplification. Alternatively, they can be chemically synthesized using one of the methods discussed in Section II.

The majority of the DNA fragments involved in protein binding are relatively short and may be synthesized in large quantities. Chemical synthesis followed by enzymic synthesis or chemical synthesis alone provides several advantages over the isolation of DNA by restriction enzyme digestion of natural DNA: (a) one can synthesize a DNA of desired length without the necessity of including any extraneous DNA; (b) one can modify the synthesis so that altered or modified bases can be introduced at specific locations and their effect on protein recognition analyzed; (c) one can obtain larger quantities of a specific DNA fragment for physicochemical studies. In this section we discuss

a few examples in which synthetic oligonucleotides have been utilized for studying protein–nucleic acid interaction.

A. The Lactose Operator–Repressor System

During the past decade, the genetics and biochemistry of the lactose operator–repressor system have been subjected to extensive study. The primary sequence of both operator and repressor has been determined. Gilbert and Maxam (151) obtained the sequence of the *lac* operator by isolating and sequencing the fragment protected by the *lac* repressor from pancreatic DNase digestion. This 27-nucleotide-long duplex fragment has a region of 2-fold symmetry, 21 nucleotides long, which was considered to be a significant feature in the specific recognition by the repressor protein.

We have recently extended the studies of this system by the use of synthetic oligonucleotides corresponding to the *lac* operator region. This section summarizes only the salient features, since a detailed account has been presented elsewhere (24). In the first step, two 21-nucleotide-long single-stranded DNAs were chemically synthesized by a modified phosphotriester method (9). The two strands were then annealed to produce a duplex 21-mer with a sequence identical to that of the *lac* operator region (Fig. 16). This duplex DNA was also synthesized by the joining of short oligonucleotide fragments with T4 DNA ligase, or by using primer extension–repair synthesis (see Section IV, A). The *in vitro* biological activity of the synthetic 21-mer duplex was examined by binding it to the *lac* repressor and assaying the complex by the Millipore filter assay. The 21-mer duplex did indeed bind to the *lac* repressor, and the binding was specific (9, 152), as shown by its drastic reduction by the inducer iPrSGal (Fig. 17). To test whether the synthetic 21-mer duplex could function as an operator *in vivo*, this sequence was introduced into bacterial plasmids by the use of suitable synthetic adaptors (see next section). When introduced into *E. coli*, the hybrid plasmids, now containing the synthetic *lac* operator, activated the β -galactosidase synthesis (15, 77, 153), which was assayed by standard procedures.

Once it was shown that the 21-nucleotides-long duplex DNA contained all the information necessary for recognition by the *lac* repressor protein *in vitro* and *in vivo*, a number of shorter duplex se-

5' AATTGTGAGCGGATAACAATT
3' TTAACACTCGCCTATTGTTAA

FIG. 16. Sequence of the chemically synthesized 21-nucleotide lactose operator DNA.

SYNTHETIC OLIGODEOXYNUCLEOTIDES

127

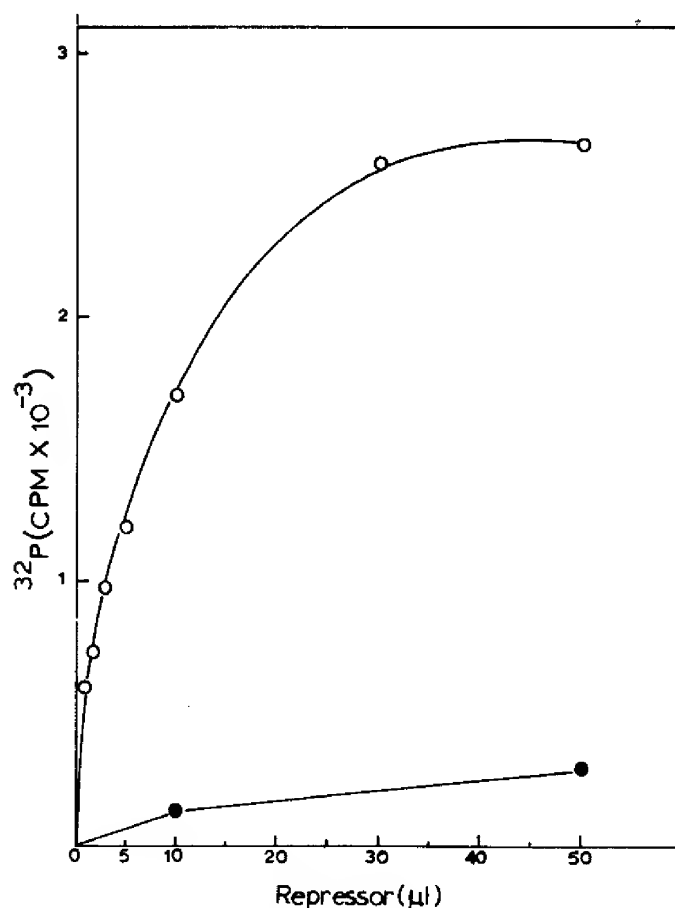


FIG. 17. Binding curve of the synthetic 21-nucleotide *lac* operator DNA to the *lac* repressor. The 5' ^{32}P labeled *lac* operator duplex was taken up in binding buffer (9), and various volumes of the *lac* repressor (15 $\mu\text{g}/\text{ml}$) were added. Triplicate samples (30 μl each) were filtered on 13-mm Millipore filters, and the filters were counted (○). Control samples contained 1 mM isopropylthiogalactoside (●) to inhibit this specific binding.

In this experiment (9), only about 10% of the input counts in the duplex operator fraction were retained on the filter, for the following reason. The single-stranded 21-mer was approximately 90% pure. In annealing to form the 21-mer duplex, 81% duplex of the correct size should be formed. However, since each single-stranded 21-mer has extensive self-complementary sequences and can form hairpins, and some nucleotides may be depurinated, the yield of completely intact 21-mer duplex was approximately 30%. In a second report (152), the binding was only 4%, owing to the fact that one of the single-stranded 21-mer segments was also contaminated with a 15-mer (the precursor of 21-mer). On the other hand, in more recent experiments (10), the 21-mer duplex was constructed by enzymic repair synthesis of synthetic 12-mer or 15-mer annealed to a 21-mer. Therefore, there is less chance for the single strands to form self-complementary structures. Furthermore, the nucleotides incorporated by repair synthesis are not damaged. As a result of these improvements, 40% retention of the input 21-mer duplex on the filter was observed in the presence of the *lac* repressor (10). With a half-life of approximately 40 seconds, 40% retention is close to the theoretical maximum.

quences were synthesized to determine the minimal essential recognition length of the *lac* operator. The shorter *lac* operator sequences were synthesized as shown in Fig. 18. In studying their binding efficiencies with the *lac* repressor, we found that a specific 17-nucleotide-long duplex *lac* operator possessed all the features necessary for recognition by the *lac* repressor (10). Finally, the 17-mer sequence was introduced into a bacterial plasmid and shown to function *in vivo* as a *lac* operator by activating β -galactosidase synthesis in *E. coli* (10).

In addition to determining the minimal essential length of the *lac* operator, several sequences with modified bases have been synthesized (9, 11, 12, 24). A detailed study of modified operators of this type should yield additional information regarding the roles of various nucleotides specifically recognized by the *lac* repressor.

B. The *E. coli* Tyrosine tRNA Gene-Promoter System

Several modified DNA sequences in the promoter region of the *E. coli* tyrosine tRNA gene have been synthesized (154) for studying the interaction of these DNA molecules with RNA polymerase. Their study has been directed toward answering the following questions. (1) Do nucleotide sequences at and following the initiation site of transcription influence RNA polymerase binding or transcription? To investigate this, the promoter is joined to parts of the synthetic gene below the natural initiation site. (2) The tyrosine tRNA gene promoter

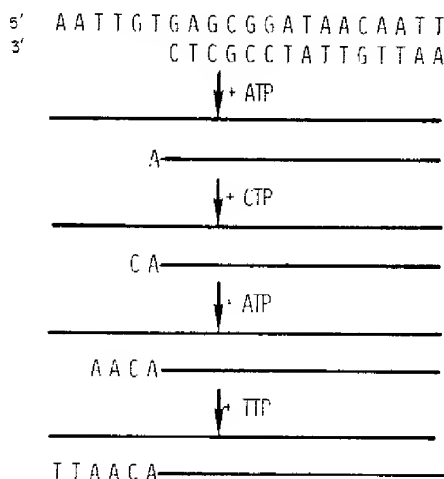


FIG. 18. The plan showing the synthesis of *lac* operator sequence of varying length by partial repair synthesis using the upper 21-mer as template and the 15-mer as primer.

shows an extremely weak binding to RNA polymerase when compared to bacteriophage promoters. Is this due to the high (G + C)-content of the tyrosine tRNA promoter region that precedes the initiation site? To test this, the structure of the promoter has been modified to lower the (G + C)-content and to remove the symmetrical sequence. Systematic studies of this kind promise to provide a wealth of information regarding the regulatory mechanism of transcription.

C. The Restriction Endonuclease System

Another area of interest in the protein-nucleic acid interaction concerns the mechanism of restriction enzyme recognition and the cleavage of DNA. Restriction endonucleases recognize DNA sequences 4 to 6 nucleotides long and cleave the DNA specifically at those sites. A self-complementary octanucleotide, d(pT-G-A-A-T-T-C-A) that contains the recognition sequence of the *EcoRI* endonuclease has been synthesized (14). This octamer can act as a substrate for both the restriction endonuclease and the methylase; the cleavage and methylation occur at the same location on the DNA. From these results, it was concluded that the substrate for the enzyme is a short double-helical segment of DNA containing the central hexamer of the octanucleotide, and that it is unlikely that any cruciform rearrangement is required for enzyme activity.

The *BamI* connector decanucleotide, d(C-C-G-G-A-T-C-C-G-G), not only contains the *BamI* site and is cleaved by *BamI* endonuclease (15), but also contains the recognition sequence d(C-C-G-G) for *HpaIII* restriction endonuclease. However, the sequences are not cleaved by *HpaII* endonuclease, presumably because the sequences are located too close to the termini of the decanucleotide. On the other hand, when two *BamI* connectors are joined by blunt-end ligation using T4 DNA ligase, the resulting 20-mer becomes an active substrate for *HpaII* (Fig. 19). The 20-mer was found to be degraded by *HpaII* endonuclease at the locations marked by the arrows to produce 5'-³²P-labeled 7-mer and 11-mer (Bahl and Wu; Smith, unpublished) and by *HaeIII* endonuclease to give the original decamer. Furthermore, the 20-mer proved to be excellent substrate for M · *HpaII* and M · *HaeIII* methylases in the presence of S-[³H]adenosylmethionine (H. Smith *et al.*, unpublished). The 20-mer can be methylated at positions 8 and 12 using M · *HpaII* methylase, or at positions 11 using M · *HaeIII* methylase.

Nucleotide recognition sequences for other restriction endonucleases, such as *HindIII* (15, 16) and *PstI* (155), have also been chemically synthesized. The *BamI* and *HindIII* endonucleases each recog-

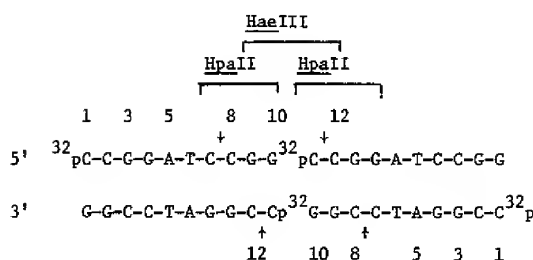


FIG. 19. Sequence of a 20-nucleotide-long duplex DNA formed by the dimerization of two 10-nucleotide-long *Bam*I connector sequence. This duplex DNA can be cleaved by restriction endonuclease *Bam*I, *Hpa*II, and *Hae*III.

nize a hexanucleotide sequence in a large DNA molecule and cleave it at that site. In synthetic decamers containing the recognition sequence for these enzymes, only the *Bam*I endonuclease produces cleavage. When either decamer is dimerized or polymerized, both endonucleases produce cleavage at the sites of recognition sequences. This indicates that the minimum size requirements for binding and cleavage in these enzymes are different. It appears that a decamer is too small for the *Hind*III endonuclease to bind, although it is large enough for the *Bam*I endonuclease to bind and cleave (Bahl, Wu and Narang, unpublished).

VII. Synthetic Oligodeoxyribonucleotides as Tools in Molecular Cloning of DNA

During the last four years, useful techniques have been developed for the *in vitro* joining of DNA segments to vehicle DNA molecules capable of independent replication (156-159). The cloning vehicle may be a plasmid DNA (156, 157), a phage λ DNA (158, 159), or an SV40 DNA (160). After joining the DNA segment to the cloning vehicle, the resulting hybrid DNA (known as chimeras) can be used to transform a suitable cell. The hybrid DNA can then be selected from among the transformed cells, and its expression in terms of DNA replication, transcription, or translation studied.

The three essential steps in cloning a DNA molecule are shown in Fig. 20. First, the DNA molecule (which contains the fragment to be cloned) is cut to a desired length. The best method is cleaving the DNA molecule with a suitable restriction enzyme (161) in such a way that only one fragment contains the sequence desired for cloning (designated as DNA-X in Fig. 20). Next, the DNA fragment is joined to the cloning vehicle DNA by means of DNA ligase. To achieve this, matching cohesive ends must be present at the termini of both DNA-X and

SYNTHETIC OLIGODEOXYNUCLEOTIDES

131

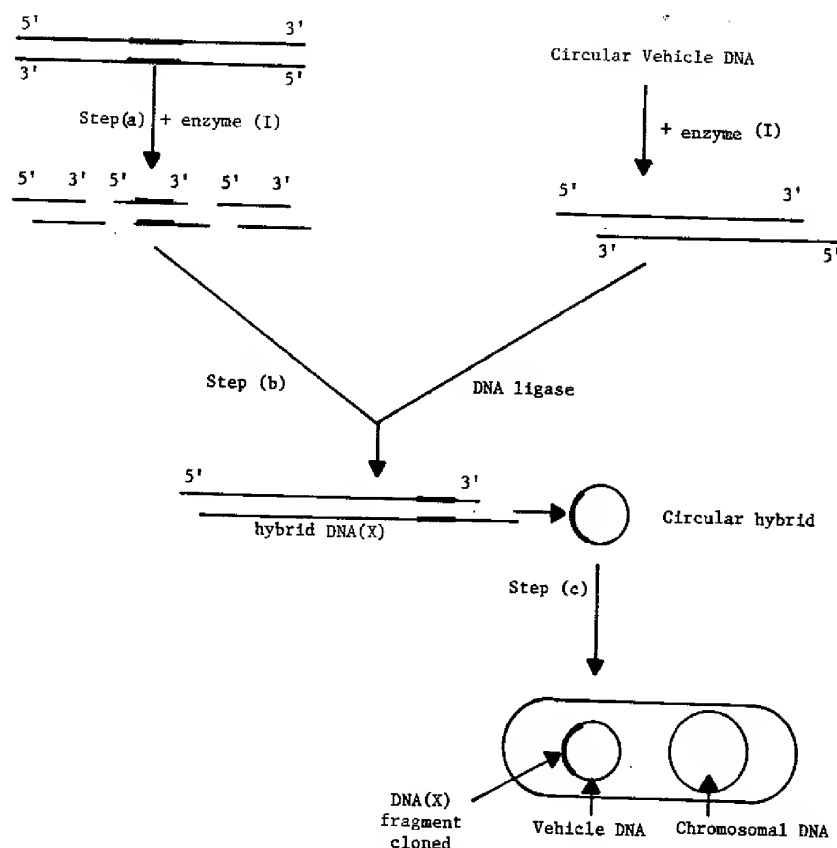


FIG. 20. A scheme for producing and joining of DNA(X) to a cloning vehicle and transforming a cell using the resulting hybrid. The segment to be cloned, designated DNA-X, is denoted by a heavy line. In step (a), both the DNA containing the segment to be cloned and the circular vehicle DNA are cut using restriction enzyme I. In step (b), DNA-X is joined to the now linear vehicle DNA using DNA ligase. In step (c), the resulting DNA-X-vehicle DNA hybrid is used to transform a permeable *E. coli* cell. Note that the circular forms represent double-stranded DNA.

the cloning vehicle. Finally, the hybrid DNA is introduced into functional living cell, and a suitable method is used for selecting the clone containing the specific hybrid DNA.

We now describe newer methods for improving the second step, the joining of DNA to the cloning vehicle, by using chemically synthesized oligodeoxynucleotides.

A. A Specific Method Involving Synthetic Cohesive Ends

A DNA segment for cloning can be chemically synthesized to include a protruding single-stranded sequence that corresponds to the recognition sequence of a restriction endonuclease. For example, *lac*

operator DNA has been synthesized (153) to include a protruding 5' d(pA-A-T-T) sequence corresponding to part of the recognition sequence of *EcoRI* restriction endonuclease (162). A synthetic *lac* operator (21-nucleotide-long duplex) containing a protruding 5' d(pA-A-T-T) sequence at each end is shown in Fig. 21. A molecule of circular pMB9 plasmid DNA has been cut once at the desired point by *EcoRI* restriction endonuclease to produce a linear pMB9 DNA with a protruding 5' d(pA-A-T-T) sequence at each end. Next, H-bonds are formed between the protruding sequences of the two types of molecules (I and II). Finally, they are joined covalently using T4 DNA ligase to produce a circular hybrid *lac*-pMB9 DNA. This hybrid is capable of transforming competent *E. coli* cells and expressing its biological activity *in vivo* (153).

Two other investigators have made independent use of the same principle. The synthetic *E. coli* tyrosine tRNA gene and its regulatory elements together with protruding 5' d(pA-A-T-T) sequences has been joined to plasmid or phage DNA for molecular cloning (40, 154). An artificial minigene coding for part of RNase, which consists of two decadeoxynucleotides containing the same protruding sequence, is being joined to a vehicle DNA for cloning (163).

In all three examples, the protruding restriction enzyme recogni-

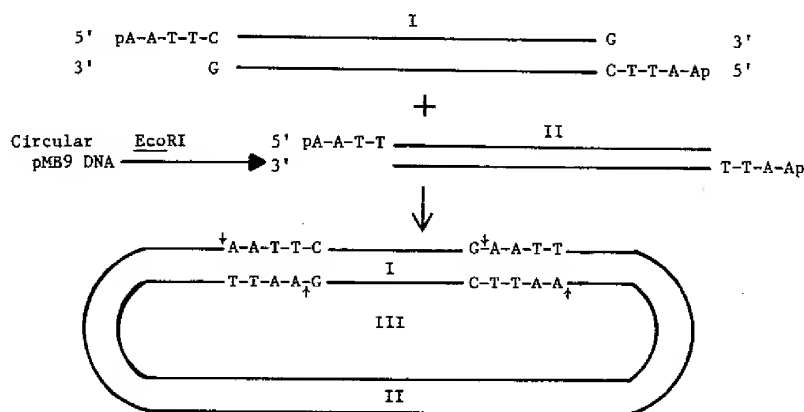


FIG. 21. Joining the chemically synthesized lactose operator duplex containing protruding 5' d(pA-A-T-T) sequence to *EcoRI* endonuclease-cut pMB9 DNA. Structure I represents the synthetic *lac* operator to be cloned, a 21-nucleotide-long duplex containing protruding 5' d(pA-A-T-T) sequences at each end. Structure II represents a linear pMB9 plasmid DNA to be used as the cloning vehicle. It is formed by cutting a circular pMB9 DNA using *EcoRI* endonuclease, and also contains a protruding 5' d(pA-A-T-T) sequence at each end. Structure III represents a circular hybrid *lac*-pMB9 DNA formed by joining structures I and II at their 5' d(pA-A-T-T) ends with T₄ DNA ligase. The arrows mark the positions where phosphodiester bonds are formed.

Board Assigned Page #2037

Fig. 22, I, *Bam*I restriction endonuclease) to create the *Bam*I endonuclease cohesive ends, 5' d(pG-A-T-C).

Alternatively, a ready-made *Bam*I adaptor (Fig. 22, II) can be used. As no digestion with *Bam*I endonuclease is required to create the cohesive end, this adaptor can be used for cloning DNA molecules that have internal *Bam*I sites. In the final step, this molecule is joined to a DNA vehicle (cut by the same endonuclease) to produce a hybrid DNA, as described earlier (see Fig. 20).

Chemically synthesized connector molecules are useful tools in cloning since the same type of oligonucleotide connector can serve to introduce a variety of double-stranded DNA molecules into cloning vehicles at specific sites. Thus far, connector molecules carrying the *Eco*RI (14, 16), *Bam*I (15, 16), *Hind*III (15, 16) and *Pst*I (155) sites have been synthesized.

C. A General Method Involving Adaptors

1. DESCRIPTION OF METHOD

A second general method has been developed in our laboratory (155). This method uses synthetic adaptors to convert one type of cohesive end to another for the cloning of DNA molecules. The principle is illustrated in Fig. 23. In many instances, a DNA molecule to be cloned (DNA-X) is most conveniently obtained by cutting a long DNA molecule with a specific restriction enzyme, and thus carrying a particular cohesive end (e.g., *Eco*RI sequence). However, the best cloning vehicle may be a DNA carrying a different cohesive end, e.g., a *Bam*I sequence. Cloning at this site of the pBR313 plasmid has the advantage of inactivating the tetracycline-resistant gene (164), thereby providing a handle for easy selection of the cloned DNA-X.

To convert one type of cohesive end to another, we synthesized

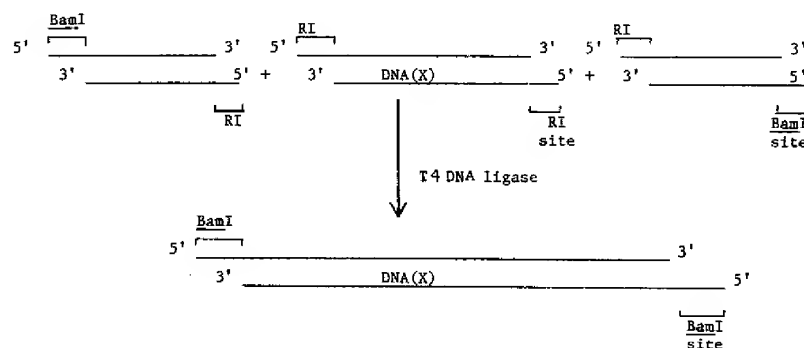


FIG. 23. A method for converting one type of cohesive end sequence to another by the use of double-headed adaptors.

NARANG

SYNTHETIC OLIGODEOXYNUCLEOTIDES

135

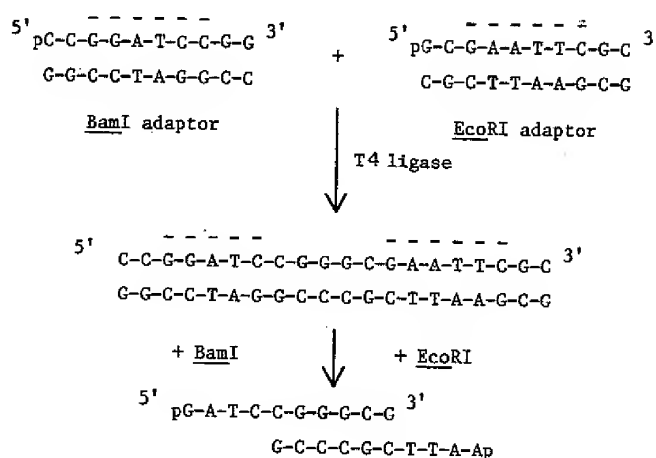


FIG. 24. A method for synthesizing a conversion adaptor by blunt-end ligation of two different types of connectors, followed by cleavage by two restriction endonucleases. Alternatively, the ready-made adaptor (lower structure) can be synthesized chemically.

several double-headed adaptor molecules possessing different cohesive ends at their termini. In Fig. 23, adaptor molecules with a *Bam*I site on one end and an *Eco*RI site on the other end are joined with DNA ligase to DNA-X by matching their *Eco*RI sites. The resulting DNA-X molecule now contains two *Bam*I sites and is ready to be joined to a cloning vehicle carrying *Bam*I sites (15).

We have synthesized several double-headed adaptors using two methods. In the first, two different even-ended connectors (such as the 10-nucleotide-long *Bam*I and *Eco*RI connectors) are joined end-to-end by blunt-end ligation as shown in Fig. 24. After purification, the desired 20-nucleotide-long double connector is cleaved by both *Bam*I and *Eco*RI endonucleases to create the double-headed adaptor. The disadvantage is that only small amounts of the adaptor can be made owing to limitation at the DNA ligase step for blunt-end ligation.

In the second method, the double-headed adaptor (such as that shown on the bottom part of Fig. 24) is directly synthesized by chemical methods. The double-headed adaptors synthesized by either method can be used for the conversion of *Eco*I sites to *Bam*I sites or vice versa. Syntheses of other double-headed adaptors, such as *Bam*I-*Hin*III and *Hin*III-*Eco*RI, have also been achieved (155).

2. CONVERSION OF 3'-PROTRUDING COHESIVE ENDS TO 5'-PROTRUDING COHESIVE ENDS BY THE USE OF SINGLE-STRANDED ADAPTORS

Several restriction enzymes cut DNA molecules to produce 3'-protruding cohesive ends, such as *Hae*II (108) and *Pst*I (158). Two

methods are available for joining a DNA molecule containing 5'-protruding ends to one containing 3'-protruding ends. In the first method, the 3' protruding ends can be removed by digestion with *Aspergillus* S₁ nuclease (165), which removes the single-strand ends completely (166). The proper connectors can then be added to the even-ended duplex DNA using the process described in Section VII, B. The disadvantage of using S₁ nuclease is that if the DNA-X molecule to be cloned contains single-strand breaks, the nuclease may produce unwanted double-strand breaks at these points. In the second method, a single-stranded decanucleotide such as 5' d(pA-A-T-T-C-A-G-C-G-C) can be ligated to the termini, thus converting 3' protruding *Hae*II sites on a DNA molecule to 5' protruding *Eco*RI sites (155).

D. Synthesis of Connectors and Adaptors of Different Lengths

After a DNA segment has been cloned, the next step is the transcription and translation of that segment. In cases where RNA synthesis is initiated from a promoter site on the vehicle DNA, correct translation requires that the reading frame be maintained, i.e., that the length of DNA-X plus the connector (or adaptor) be a multiple of 3. Since in most cases the exact length of DNA-X is unknown, connectors of 3 different lengths that will add n , $n + 1$ or $n - 1$ nucleotides (where n is a multiple of 3) must be synthesized. For example, ligation of a linker molecule B' (Fig. 22) to a DNA will add 6 nucleotides ($n + 0$). For adding $n + 1$ nucleotides, the hexanucleotide double-stranded portion of linker B' can be changed to a heptanucleotide. Connectors of this type are being synthesized in our laboratory.

VIII. Concluding Remarks

The techniques developed for the synthesis of oligonucleotides of defined sequences can be used as primers for DNA sequence analysis. They can also be applied to produce modified genes or genetic elements. These, in turn, may be used to determine which parts of a molecule under consideration are essential to its functioning and which are not. For example, the lactose operator may be synthesized with altered bases or base-pairs at prescribed locations. The effects of these alterations, whether they be an increase, decrease, or no change in the binding of lactose repressor, may then be studied. Similarly, alterations of one or more base-pairs of a promoter site may be made, and the modified promoter tested for interaction with RNA polymerase. Once the essential features for promoter-RNA polymerase interaction are thoroughly understood, it may be possible to design and synthesize a sequence with enhanced promoter activity. Such a "super" promoter could then be joined to a gene coding for a

desired polypeptide, and that in turn could be joined to a cloning vehicle to produce a very efficient system for transcription of this particular gene *in vivo*. Optimal conditions may be found so that large quantities of this RNA transcript would then be used to direct the synthesis in the microorganism of this protein product in bulk quantities. The potential for large-scale production of scarce biologically or medically useful products could have far-reaching benefits.

ACKNOWLEDGMENTS

We are very grateful to our friends who kindly provided us with manuscripts in advance of publication. We thank Drs. R. Brousseau, K. Itakura, T. Hozumi, E. Jay, N. Katagiri, K. J. Marians, J. J. Michiniwicz, R. Roychoudhury, A. K. Sood, J. Stawinsky, J. Szostak, and D. Tu for participation in some of the experiments described here. Research that originated in the authors' laboratories was supported by research grants CA-14989 and GM-18887 from the National Institutes of Health, Grant PCM-76-05072 from the National Science Foundation and Grant 16124 from the National Research Council of Canada.

REFERENCES

1. A. Kornberg, "DNA Synthesis." Freeman, San Francisco, 1974.
2. H. G. Khorana, *Harvey Lect. Ser.* **62**, 79 (1968).
3. R. Lohrmann, D. Söll, H. Hayatsu, E. Ohtsuka and H. G. Khorana, *JACS* **88**, 819 (1966).
4. K. L. Agarwal, H. Büchi, M. H. Caruthers, N. Gupta, H. G. Khorana, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. RajBhandary, J. H. van de Sande, V. Sgaramella, H. Weber and T. Yamada, *Nature* **227**, 27 (1970).
5. H. G. Khorana, K. L. Agarwal, P. Besmer, H. Büchi, M. H. Caruthers, P. J. Cashion, M. Fridkin, E. Jay, K. Kleppe, R. Kleppe, A. Kumar, P. C. Loewen, R. C. Miller, K. Minamoto, A. Panet, U. L. RajBhandary, B. Ramamoorthy, T. Sekiya, T. Takeya and J. H. van de Sande, *JBC* **251** 565 (1976).
6. C. L. Harvey, K. Olson, A. de Czekala and A. L. Nussbaum, *Nucl. Acids Res.* **2**, 2007 (1975).
7. H. Köster, H. Blöcker, R. Frank, G. Geussenhainer and W. Kaiser, *Hoope Seylers Z. Physiol. Chem.* **356**, 1585 (1975).
8. K. Itakura, N. Katagiri, S. A. Narang, C. P. Bahl, K. J. Marians and R. Wu, *JBC* **250**, 4592 (1975).
9. C. P. Bahl, R. Wu, K. Itakura, N. Katagiri and S. A. Narang, *PNAS* **73**, 91 (1976).
10. C. P. Bahl, R. Wu, J. Stawinsky and S. A. Narang, *PNAS* **74**, 966 (1977).
11. J. Stawinski, T. Hozumi, S. A. Narang, C. P. Bahl and R. Wu, *Nucl. Acids Res.* **4**, 353 (1977).
12. D. V. Goeddel, D. G. Yansura and M. H. Caruthers, *Nucl. Acids Res.* **4**, 3039 (1977).
13. D. G. Yansura, D. V. Goeddel and M. H. Caruthers, *Bchem* **16**, 1772 (1977).
14. P. J. Greene, M. S. Poonian, A. L. Nussbaum, L. Tobias, D. E. Garfin, H. W. Boyer and H. M. Goodman, *JMB* **99**, 237 (1975).
15. C. P. Bahl, K. J. Marians, R. Wu, J. Stawinski and S. A. Narang, *Gene* **1**, 81 (1976).
16. R. H. Scheller, R. E. Dickerson, H. W. Boyer, A. D. Riggs, and K. Itakura, *Science* **196**, 177 (1977).
17. R. Wu, R. Bambara and E. Jay, *CRC Crit. Rev. Biochem.* **2**, 455 (1974).

18. H. Kössel and H. Seliger, *Progr. Chem. Org. Nat. Prod.* **32**, 298 (1975).
19. S. A. Narang and R. H. Wightman, in "Total Synthesis of Natural Products" (J. W. Ap Simon, ed.), Vol. 1, p. 280. Wiley (Interscience), New York, 1973.
20. V. Amarnath and A. D. Broom, *Chem. Rev.* **77**, 183 (1977).
21. K. L. Agarwal, A. Yamazaki, P. J. Cashion and H. G. Khorana, *Angew. Chem.* **11**, 451 (1972).
22. H. G. Khorana, in "Reflections on Biochemistry" (A. Kornberg, ed.), p. 273. Pergamon, Oxford, 1976.
23. A. L. Nussbaum, *Biorgan. Chem.* **3**, 278 (1970).
24. R. Wu, C. P. Bahl and S. A. Narang, *Curr. Top. Cell. Regul.* **13**, in press (1977).
25. R. Wu, E. Jay and R. Roychoudhury, in "Methods in Cancer Research" (H. Busch, ed.), Vol. 12, p. 87. Academic Press, New York, 1976.
26. R. D. Wells, R. W. Blakesley, S. C. Hardies, G. T. Horn, J. E. Larson, E. Selsing, J. F. Burd, H. W. Chan, J. B. Dodgson, K. F. Jensen, I. F. Nes and R. M. Wartell, *CRC Crit. Rev. Biochem.* **5**, 305 (1977).
27. S. A. Narang, C. P. Bahl and R. Wu, *Can. J. Biochem.* **55**, 1125 (1977).
28. A. M. Michelson and A. R. Todd, *JCS* 2632 (1955).
29. P. T. Gilham and H. G. Khorana, *JACS* **80**, 6212 (1958).
30. T. M. Jacob and H. G. Khorana, *JACS* **86**, 1630 (1964).
31. R. Lohrmann and H. G. Khorana, *JACS* **88**, 829 (1966).
32. R. K. Ralph and H. G. Khorana, *JACS* **83**, 2926 (1961).
33. S. A. Narang, T. M. Jacob and H. G. Khorana, *JACS* **89**, 2158 (1967).
34. E. Ohtsuka and H. G. Khorana, *JACS* **89**, 2195 (1967).
35. S. A. Narang, K. Itakura, C. P. Bahl and Y. Y. Wigfield, *BBRC* **49**, 445 (1972).
36. S. A. Narang, K. Itakura, C. P. Bahl and N. Katagiri, *JACS* **96**, 7074 (1974).
37. M. T. Doel and M. Smith, *FEBS Lett.* **34**, 99 (1973).
38. R. Padmanabhan, E. Jay and R. Wu, *PNAS* **71**, 2510 (1974).
39. R. Wu, C. D. Tu and R. Padmanabhan, *BBRC* **55**, 1092 (1973).
40. R. Belagaje, E. L. Brown, H. J. Fritz, M. J. Gait, R. G. Lees, K. Norris, T. Sekiya, T. Takeya, R. Contreras, H. Küpper, M. J. Ryan and H. J. Khorana, *ACS Adv. Chem. Ser.* in press (1977).
41. D. G. Knorre and V. F. Zarytova *Nucl. Acids Res.* **3**, 2709 (1976).
42. S. A. Narang, O. S. Bhanot, J. Goodchild, R. H. Wightman and S. K. Dheer, *JACS* **94**, 6183 (1972).
43. K. L. Agarwal, Y. A. Berlin, H. J. Fritz, M. J. Gait, D. G. Kleid, R. C. Lees, K. E. Norris, B. Ramamoorthy and H. G. Khorana, *JACS* **98**, 1065 (1976).
44. T. F. Gabriel and J. Michalewsky, *J. Chromatogr.* **67**, 309 (1972).
45. H. J. Fritz, R. Belagaje, E. L. Brown, R. H. Fritz, R. A. Jones, R. C. Lees and H. G. Khorana, *Bchem.* in press (1978).
46. M. J. Gait and R. C. Sheppard, *JACS* **98**, 8514 (1976).
47. K. L. Agarwal and H. G. Khorana, *JACS* **94**, 3578 (1972).
48. R. L. Letsinger and K. K. Ogilvie, *JACS* **89**, 4801 (1967).
49. R. L. Letsinger and K. K. Ogilvie, *JACS* **91**, 3350 (1969).
50. N. J. Cusack, C. B. Reese and J. H. van Boom, *Tetrahedron Lett.* 2209 (1973).
51. F. Eckstein and I. Risk, *Chem. Ber.* **102**, 2362 (1969).
52. T. Neilson and E. S. Werstiuk, *Can. J. Chem.* **49**, 3004 (1971).
53. J. Smrt, *Coll. Czech. Chem. Commun.* **39**, 972 (1974).
54. M. Sekine and T. Hata, *Tetrahedron Lett.* 1711 (1975).
55. K. Itakura, C. P. Bahl, N. Katagiri, J. J. Michniewicz, R. H. Wightman and S. A. Narang, *Can. J. Chem.* **51**, 3649 (1973).
56. J. C. Catlin and F. Cramer, *J. Org. Chem.* **38**, 245 (1973).

57. N. Katagiri, K. Itakura and S. A. Narang, *JACS* **97**, 7332 (1975).
58. N. Katagiri, K. Itakura and S. A. Narang, *Chem. Commun.* 326 (1974).
59. K. Itakura, N. Katagiri, C. P. Bahl, R. H. Wightman and S. A. Narang, *JACS* **97**, 7327 (1975).
60. J. Stawinski, T. Hozumi and S. A. Narang, *Can. J. Chem.* **54**, 670 (1976).
61. Y. A. Berlin, O. G. Chakmakheva, V. A. Efimov, M. N. Kolosov and Y. G. Koobko, *Tetrahedron Lett.* 1353 (1973).
62. J. H. van Boom and P. M. J. Burgers, *Tetrahedron Lett.* 4875 (1976).
63. A. K. Sood and S. A. Narang, *Nucl. Acids Res.* **4**, 2757 (1977).
64. R. W. Adamiak, M. Z. Barciszewska, E. Biala, K. Grzetskowiak, R. Kierzek, A. Kraszewski, W. T. Markiewicz and M. Wiewirowski, *Nucl. Acids Res.* **3**, 3397 (1976).
65. S. A. Narang and R. Wu, "Practice and Applications of Thin Layer Densitometry" (J. C. Touchstone, ed.), in press. Wiley, New York, 1978.
66. R. L. Letsinger and W. B. Lunsford, *JACS* **98**, 3655 (1976).
67. M. Cellert, *PNAS* **57**, 148 (1967).
68. B. M. Olivera and I. R. Lehman, *PNAS* **57**, 1426 (1967).
69. M. L. Gefter, A. Becker and J. Hurwitz, *PNAS* **58**, 240 (1967).
70. N. K. Gupta, E. Ohtsuka, H. Weber, S. H. Chang and H. G. Khorana, *PNAS* **60**, 285 (1968).
71. N. K. Gupta and H. G. Khorana, *PNAS* **61**, 215 (1968).
72. H. G. Khorana, *JMB* **72**, 209 (1972).
73. V. Sgaramella, J. H. van de Sande and H. G. Khorana, *PNAS* **67**, 1468 (1970).
74. V. Sgaramella, *PNAS* **69**, 3389 (1972).
75. A. Sugino, H. M. Goodman, H. L. Heyneker, J. Shine, H. W. Boyer and N. R. Cozzarelli, *JBC* **252**, 3987 (1977).
76. K. Backman, M. Ptashne, W. Gilbert, *PNAS* **73**, 4174 (1976).
77. H. L. Heyneker, J. Shine, H. M. Goodman, H. W. Boyer, J. Rosenberg, R. E. Dickerson, S. A. Narang, K. Itakura, S. Lin and A. Riggs, *Nature* **263**, 748 (1976).
78. C. C. Richardson, R. B. Inman and A. Kornberg, *JMB* **9**, 46 (1964).
79. R. Wu and A. D. Kaiser, *JMB* **35**, 523 (1968).
80. R. Wu, *JMB* **51**, 501 (1970).
81. R. Wu, J. Donelson, R. Padmanabhan and R. Hamilton, *Bull. Inst. Pasteur* **70**, 203 (1972).
82. A. C. Lezius and K. H. Scheit, *EJB* **3**, 85 (1967).
83. R. Silber, V. G. Malathi and J. Hurwitz, *PNAS* **69**, 3009 (1972).
84. T. J. Snopek, A. Sugino, K. L. Agarwal and N. R. Cozzarelli, *BBRC* **68**, 417 (1976).
85. S. Gillam and M. Smith, *Nature NB* **238**, 233 (1972).
86. S. Gillam, K. Waterman, M. Doel and M. Smith, *Nucl. Acids Res.* **1**, 1649 (1974).
87. S. Gillam, F. Rottman, P. Jahnke and M. Smith, *PNAS* **74**, 96 (1977).
88. H. Kössel and R. Roychoudhury, *EJB* **22**, 271 (1971).
89. R. Roychoudhury and H. Kössel, *EJB* **22**, 310 (1971).
90. T. Sekiya and H. G. Khorana, *PNAS* **71**, 2978 (1974).
91. R. Padmanabhan, R. Wu and R. Calendar, *JBC* **249**, 6197 (1974).
92. R. Padmanabhan, *Bchem.* **16**, 1996 (1977).
93. R. Roychoudhury and R. Wu, *JBC* **252**, 4786 (1977).
94. C. C. Richardson, *PNAS* **54**, 158 (1965).
95. A. Novogrodsky and J. Hurwitz, *JBC* **241**, 2923 (1966).
96. B. Weiss and C. C. Richardson, *JMB* **23**, 405 (1967).
97. R. Wu and A. D. Kaiser, *PNAS* **57**, 170 (1967).
98. V. Ling, *JMB* **64**, 87 (1972).

99. K. Murray and N. E. Murray, *Nature NB* **243**, 134 (1973).
100. C. D. Tu, E. Jay, C. P. Bahl and R. Wu, *Anal. Biochem.* **74**, 73 (1976).
101. A. Maxam and W. Gilbert, *PNAS* **74**, 560 (1977).
102. R. Wu and E. Taylor, *JMB* **57**, 491 (1971).
103. P. T. Englund, *JMB* **66**, 209 (1972).
104. P. H. Weigel, P. T. Englund, K. Murray and R. W. Old, *PNAS* **70**, 1151 (1973).
105. G. S. Changas, E. Jay, R. Bambara and R. Wu, *BBRC* **54**, 998 (1973).
106. H. Kössel, R. Roychoudhury, D. Fischer and A. Otto, in "Methods in Enzymology," Vol. 29 (L. Grossman and K. Moldave, eds.), p. 322. Academic Press, New York, 1974.
107. R. Roychoudhury, E. Jay and R. Wu, *Nucl. Acids Res.* **3**, 863 (1976).
108. C. D. Tu, R. Roychoudhury and R. Wu, *BBRC* **72**, 355 (1976).
109. R. Wu, *Nature NB* **236**, 198 (1972).
110. F. Sanger, J. E. Donelson, A. R. Coulson, H. Kössel and D. Fischer, *PNAS* **70**, 1209 (1973).
111. P. C. Loewen and H. G. Khorana, *JBC* **248**, 3489 (1973).
112. P. C. Loewen, T. Sekiya and H. G. Khorana, *JBC* **249**, 217 (1974).
113. F. Sanger and A. R. Coulson, *JMB* **94**, 441 (1975).
114. H. Schott, D. Fisher and H. Kössel, *Bchem* **12**, 3447 (1973).
115. P. Besmer, R. C. Miller, Jr., M. H. Caruthers, A. Kumar, K. Minamoto, J. H. van de Sande, N. Sidarova and H. G. Khorana, *JMB* **72**, 503 (1972).
116. H. D. Robertson, B. G. Barrell, H. L. Weith and J. E. Donelson, *Nature NB* **241**, 38 (1973).
117. J. W. Szostak, J. E. Stiles, C. P. Bahl and R. Wu, *Nature* **265**, 61 (1977).
118. E. P. Heimer, M. Ahmad and A. L. Nussbaum, *BBRC* **48**, 348 (1972).
119. R. Bambara and R. Wu, *JBC* **250**, 4607 (1975).
120. T. Maniatis, M. Ptashne, B. G. Barrell and J. Donelson, *Nature* **250**, 394 (1974).
121. B. G. Barrell, G. M. Air and C. A. Hutchison III, *Nature* **264**, 34 (1976).
122. G. M. Air, E. H. Blackburn, A. R. Coulson, F. Galibert, F. Sanger, J. W. Sedat and E. B. Ziff, *JMB* **107**, 445 (1976).
123. F. Sanger, G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison III, P. M. Slocombe and M. Smith, *Nature* **265**, 687 (1977).
124. N. L. Brown and M. Smith, *Nature* **265**, 695 (1977).
125. W. Szybalski, H. Kubinski, Z. Hradecna and W. C. Summers, in "Methods in Enzymology," Vol. 21 (L. Grossman and K. Moldave, eds.), p. 383. Academic Press, New York, 1971.
126. P. T. Englund, *JBC* **246**, 3269 (1971).
127. F. Sanger, G. G. Brownlee and B. G. Barrell, *JMB* **13**, 373 (1965).
128. B. G. Barrell *Proc. Nucl. Acids* **2**, 751 (1971).
129. G. G. Brownlee and F. Sanger, *EJB* **11**, 395 (1969).
130. E. Jay and R. Wu, *Bchem* **15**, 3612 (1976).
131. E. Jay, R. Bambara, R. Padmanabhan and R. Wu, *Nucl. Acids Res.* **1**, 331 (1974).
132. R. Bambara, E. Jay and R. Wu, *Nucl. Acids Res.* **1**, 1503 (1974).
- 132a. E. Jay, R. Roychoudhury and R. Wu, *BBRC* **69**, 678 (1976).
133. R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick and A. Zamir, *Science* **147**, 1462 (1965).
134. R. Calendar and P. Berg, *Bchem* **5**, 1681 (1966).
135. H. M. Goodman, J. Abelson, A. Landy, S. Brenner and J. D. Smith, *Nature* **217**, 1019 (1968).
136. H. M. Goodman, J. N. Abelson, A. Landy, S. Zadrazil and J. D. Smith, *EJB* **13**, 461 (1970).

NARANG

SYNTHETIC OLIGODEOXYNUCLEOTIDES

141

137. F. Harada, H. J. Cross, F. Kimura, S. H. Chang, S. Nishimura and U. L. RajBhandary, *BBRC* **33**, 299 (1968).
138. F. Harada, F. Kumura and S. Nishimura, *FP* **28**, 409 (1969).
139. J. D. Smith, L. Barnett, S. Brenner and R. L. Russel, *JMB* **54**, 1 (1970).
140. J. D. Smith, K. Anderson, A. Cashmore, M. L. Hooper and R. L. Russel *CSHSQB* **35**, 21 (1970).
141. S. Altman, S. Brenner and J. D. Smith, *JMB* **56**, 195 (1971).
142. S. Altman and J. D. Smith, *Nature NB* **233**, 35 (1971).
143. T. Sekiya, R. Contreras, H. Küpper, A. Landy and H. G. Khorana, *JBC* **251**, 5124 (1976).
144. M. J. Ryan, R. Belagaje, E. L. Brown, H-J. Fritz, R. H. Fritz, M. J. Gait, H. Küpper, R. G. Lees, K. Norris, T. Sekiya and T. Takeya, *FP* **36**, 732 (1977).
145. S. A. Narang and S. K. Dheer, *Bchem* **8**, 3443 (1969).
146. E. M. Southern, *JMB* **98**, 503 (1975).
147. P. Cashion, H. Notman, G. Sathe, T. Cadger, K. Porter and E. Jay, *Nucl. Acids Res.* **4**, 2593 (1977).
148. A. Efstratiadis, F. C. Kafatos, A. Maxam and T. Maniatis, *Cell* **7**, 279 (1976).
149. R. Higuchi, G. V. Paddock, R. Wall and W. Salser, *PNAS* **73**, 3146 (1976).
150. M. Grunstein and D. S. Hogness, *PNAS* **72**, 3961 (1975).
151. W. Gilbert and A. Maxam, *PNAS* **70**, 3581 (1973).
152. S. Lin, K. Itakura, J. M. Rosenberg, G. Wilcox, C. Bahl, R. Wu, S. A. Narang, R. Dickerson and A. D. Riggs, in "Molecular Mechanisms in the Control of Gene Expression" (D. P. Nierlich, *et al.*, eds.) Vol. 5, p. 143. Academic Press, New York, 1976.
153. K. J. Marians, R. Wu, J. Stawinski, T. Hozumi and S. A. Narang, *Nature* **263**, 744 (1976).
154. E. L. Brown, R. Belagaje, H. J. Fritz, R. H. Fritz and K. Norris, *FP* **36**, 732 (1977).
155. C. P. Bahl, R. Wu, R. Brousseau, A. K. Sood, H. M. Hsiung and S. A. Narang, *BBRC* in press (1978).
156. S. N. Cohen, A. C. Y. Chang, H. W. Boyer and R. B. Helling, *PNAS* **70**, 3240 (1973).
157. V. Hershfield, H. W. Boyer, M. A. Lovett, C. Yanofsky and D. R. Helinski, *PNAS* **71**, 3455 (1974).
158. M. Thomas, J. R. Cameron and R. W. Davis, *PNAS* **71**, 4579 (1974).
159. N. E. Murray and K. Murray, *Nature* **251**, 476 (1974).
160. A. L. Nussbaum, D. Ganem, D. Davoli and G. C. Fareed, *PNAS* **73**, 1068 (1976).
161. R. J. Roberts, *CRC Crit. Rev. Biochem.* **4**, 123 (1976).
162. J. Hedgpeth, H. M. Goodman and H. W. Boyer, *PNAS* **69**, 3448 (1972).
163. M. S. Poonian, W. W. McComas and A. L. Nussbaum, *Gene* **1**, 357 (1977).
164. F. Bolivar, R. L. Rodriguez, M. C. Betlach and H. W. Boyer, *Gene* **2**, 75 (1977).
165. V. M. Vogt, *EJB* **33**, 192 (1973).
166. G. S. Ghangas and R. Wu, *JBC* **250**, 4601 (1975).

TAB II

A Sensitive Radioimmunoassay for Detecting Products Translated from Cloned DNA Fragments

Henry A. Erlich,* Stanley N. Cohen†
and Hugh O. McDevitt*

* Division of Immunology
Department of Medicine

† Departments of Medicine and Genetics
Stanford University School of Medicine
Stanford, California 94305

Summary

A simple and sensitive radioimmunoassay using *E. coli* β -galactosidase as a model protein has been developed for the detection of specific translation products of foreign gene fragments cloned into plasmid or phage vectors. This immunoassay is based upon the coupling to an insoluble matrix of $F(ab)_2$ fragments derived from the specific antiserum by pepsin digestion. The in situ analysis of phage plaques or of bacterial colonies is performed by overlaying the phage plaques or lysed bacterial colonies with a cellulose filter to which $F(ab)_2$ fragments have been chemically coupled. The antigen bound to the filter is detected by subsequent incubations with undigested antiserum and with ^{125}I -labeled *Staphylococcus aureus* protein A followed by autoradiography. By coupling the $F(ab)_2$ fragments to the wells of a plastic microtiter plate, liquid cultures can be analyzed quantitatively for the presence of antigen, making possible the analysis of heterogeneous cultures by sib selection. The detection threshold of the microtiter plate assay for liquid culture is shown to be $<2 \times 10^6$ molecules, or about 1 molecule of β -galactosidase per cell. The in situ immunoassay for bacterial colonies, which permits examination of about 1000 clones per plate, can easily detect microcolonies producing about 10 molecules of β -galactosidase per cell, while the in situ phage plaque assay, also capable of screening about 1000 plaques per plate, is even more sensitive, detecting $<1 \times 10^7$ molecules per bacteriophage plaque.

Introduction

The cloning of specific eucaryotic DNA sequences in *E. coli* requires a method for identifying those clones that have acquired plasmids or bacteriophages containing the desired DNA fragments. In some instances, the biological activity of a gene product encoded by the cloned DNA fragment can be used for the direct phenotypic selection of clones in which the specific DNA sequence is expressed as a functional protein (Struhl, Cameron and Davis, 1976; Ratzkin and Carbon, 1977; Vap-

nek et al., 1977). Such detection systems, however, require a homologous bacterial function that can be inactivated in the host strain and, consequently, are usually limited to those eucaryotic proteins which have a bacterial counterpart.

Identification of bacterial clones that contain specific eucaryotic genes inserted into plasmid or bacteriophage vectors have also utilized a messenger RNA (mRNA) or complementary DNA (cDNA) probe. Such probes have been used for in situ nucleic acid hybridization (Grunstein and Hogness, 1975; Jones and Murray, 1975) and for filter hybridization in a sib selection system (Kedes et al., 1975). Detection of this kind, however, is dependent upon the availability of highly purified nucleic acid probes.

Structural analysis of restriction endonuclease-generated DNA fragments (Morrow et al., 1974) and direct DNA sequence analysis (Ulrich et al., 1977) have been used to detect and identify eucaryotic DNA segments contained in plasmid molecules isolated from transformed *E. coli* cells. Analysis of bacterial cell extracts for the appearance of new protein bands in polyacrylamide gels has also been employed (Chang et al., 1975; Meagher et al., 1977), but such methods appear to be more suited to the analysis of specific clones than to the mass screening of bacterial populations.

Perhaps the most general assay for screening bacterial clones for gene products encoded by cloned foreign DNA is one based on antibody-mediated recognition of antigenic determinants. (This approach requires that the determinants to be recognized be protein and not carbohydrate in nature.) Such an immunoassay would potentially be capable of detecting incompletely translated products as well as proteins which have no easily detectable or selectable function. Furthermore, proteins that are unable to function in a bacterial cell environment might nevertheless retain their immunologically reactive sites. Recently, in situ immunoassays have been reported, based on enzyme-coupled antibody detection of protein in λ plaques (Sanzey et al., 1976) and on immunoprecipitation in top-layer agar for λ plaques and bacterial colonies (Skalka and Shapiro, 1976). Such assays, however, have been limited in their sensitivity. In experiments in which the expression of cloned eucaryotic DNA in procaryotic cells has been demonstrated (Struhl et al., 1976; Ratzkin and Carbon, 1977; Vapnek et al., 1977), the evidence suggests a low level of phenotypic expression. Consequently, sensitivity is a primary requirement for an experimentally useful system for detecting specific clones.

We report a novel indirect radioimmunoassay using *E. coli* β -galactosidase as a model gene

product. This assay is dependent upon the coupling of $F(ab)'_2$ fragments (bivalent immunoglobulin G fragments lacking the Fc portion) to an insoluble matrix. Such $F(ab)'_2$ fragments provide a means of specifically binding antigen to the solid phase, yet are unrecognized by the detecting reagent, ^{125}I -labeled Staph. aureus protein A, due to the absence of the Fc segment. The indirect radioimmunoassay therefore does not require purification and labeling of specific antibody. It is highly sensitive and quantitative, and can be used for screening individual clones or for sib selection analysis of heterogeneous cultures. In addition, it has been adapted for the in situ analysis of bacterial colonies and phage plaques, and can thus be used for both plasmid and phage vector cloning systems.

Results

The initial reaction in the multi-step radioimmunoassay is the coupling of $F(ab)'_2$ fragments derived from pepsin digestion of rabbit anti-E. coli β -galactosidase to an insoluble matrix. After the initial coupling step, the solid phase $F(ab)'_2$ fragments are incubated with either soluble β -galactosidase or cell lysates. Next, the undigested antiserum is allowed to interact with the antigen which is bound, via the $F(ab)'_2$ fragment, to the solid phase. In the final step, ^{125}I -labeled protein A (from Staph. aureus), which specifically recognizes the immunoglobulin Fc portion (Forsgren and Sjoquist, 1966), binds to the bound immunoglobulin. After each step, unbound reagents are removed by washing. The binding sequence is represented diagrammatically in Figure 1, and the dependence of final ^{125}I binding upon each previous binding reaction is presented in Table 1. The wells of a polyvinyl chloride microtiter plate were used as the insoluble matrix for sib selection assays of bacteria growing in liquid culture; for the in situ colony or phage plaque assay, we used a chemically activated cellulose filter. In the microtiter plate assay, quantitation is obtained by cutting out the flexible wells and measuring the bound ^{125}I counts. For mass screening, the microtiter plates are analyzed by autoradiography.

Plate Assay

The sensitivity of the indirect immunoassay can be

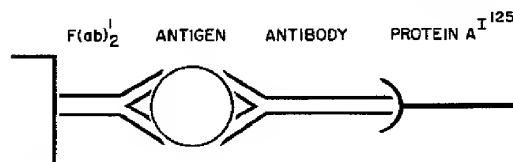


Figure 1. Schematic Diagram of the Indirect Radioimmunoassay

estimated with graded amounts of soluble β -galactosidase (Figure 2A). The detection threshold is below 0.2 ng for the pure protein. The same assay monitored by autoradiography is shown in Figure 2B. In this autoradiogram, the lower limit of detection is about 0.5 ng.

For linear quantitation of antigen, all reagents [$F(ab)'_2$, antiserum and protein A ^{125}I] must be present at saturating levels. The background level of nonspecific binding, however, rises slightly as the concentration of the detecting reagents, particularly antiserum, is increased. The concentration of these reagents should therefore be varied according to the estimated amounts of antigen to be detected. Thus the signal to noise ratio in the data presented in Table 1 can be significantly improved by using higher concentrations of antiserum, although the background radioactivity will rise slightly. The antiserum concentration in Table 1 was optimized for detecting minute amounts of β -galactosidase (Figure 2). Under the conditions of the assay shown in Figure 2, antigen is limiting only below 1 ng.

For the detection of β -galactosidase in bacterial cells, cultures of MC 1000 (lac^+) and D7001 (lac^+) were grown to saturation in L broth. For mass screening of individual clones, microcultures (300 μ l) can be grown from single colonies or from small inocula for sib selection in the wells of microtiter plates. Bacterial cells were lysed by adding 2 drops of chloroform per ml culture and SDS to a final concentration of 0.1%. The lysates (50 μ l) were added to the $F(ab)'_2$ -coated wells and reacted in the multi-step assay. The results from a typical experiment are shown in Figure 3.

The initial rise in bound radioactivity as the lysate is diluted (that is, a "prozone" effect) suggests that

Table 1. Binding of Protein A ^{125}I as a Function of Previous Binding Reactions in a Microtiter Plate

| $F(ab)'_2$ | Antigen | Antiserum | Bound ^{125}I cpm (Protein A) |
|------------|---------|-----------|------------------------------------|
| | | | SE |
| + | + | + | 12,059 \pm 139 |
| + | + | - | 268 \pm 13 |
| + | - | + | 1,109 \pm 40 |
| - | + | + | 274 \pm 8 |

The wells of a polyvinyl chloride microtiter plate were coated with $F(ab)'_2$ fragments (75 μ l of 0.025 mg/ml) or 1% BSA (50 μ l). Samples (50 μ l) of β -galactosidase solution (1 μ g/ml) or of 1% BSA (50 μ l) were incubated in the coated wells, followed by an incubation with 50 μ l of anti- β -galactosidase serum (diluted 1/2000) or of normal rabbit serum (diluted 1/2000). Finally, 50 μ l of ^{125}I -labeled protein A were added to the wells, and the plate was treated as described in Experimental Procedures. All samples were tested in quadruplicate.

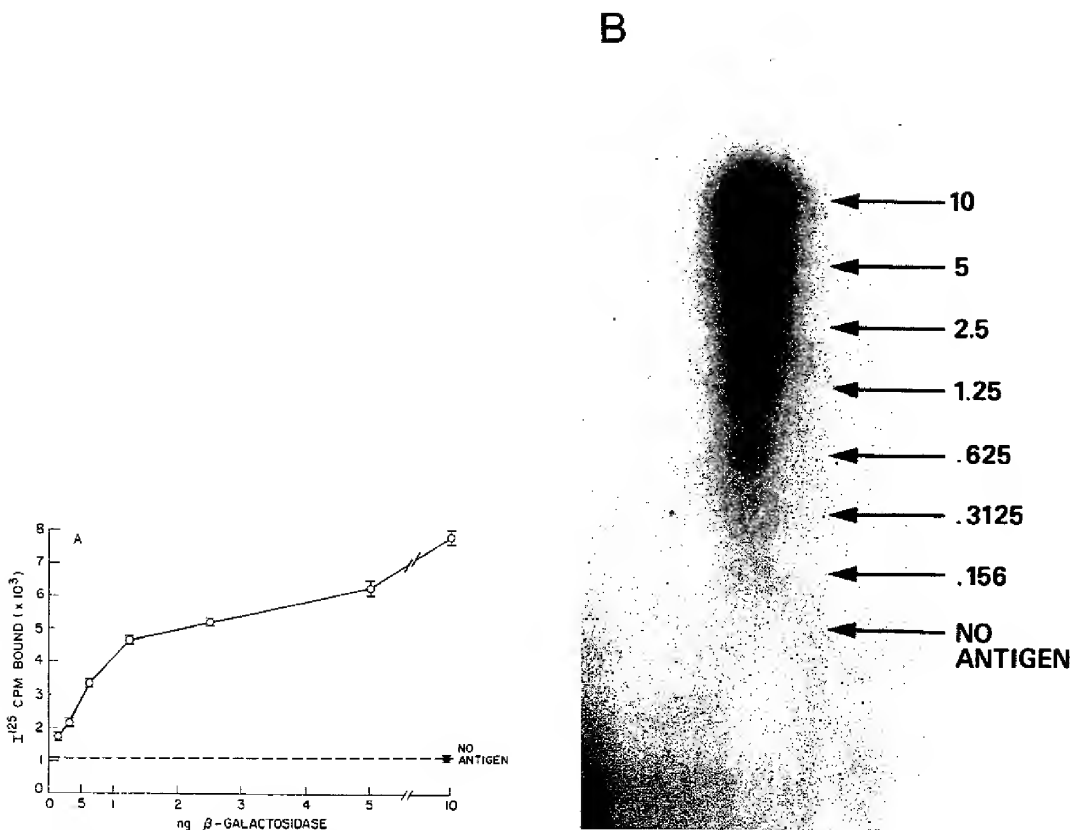


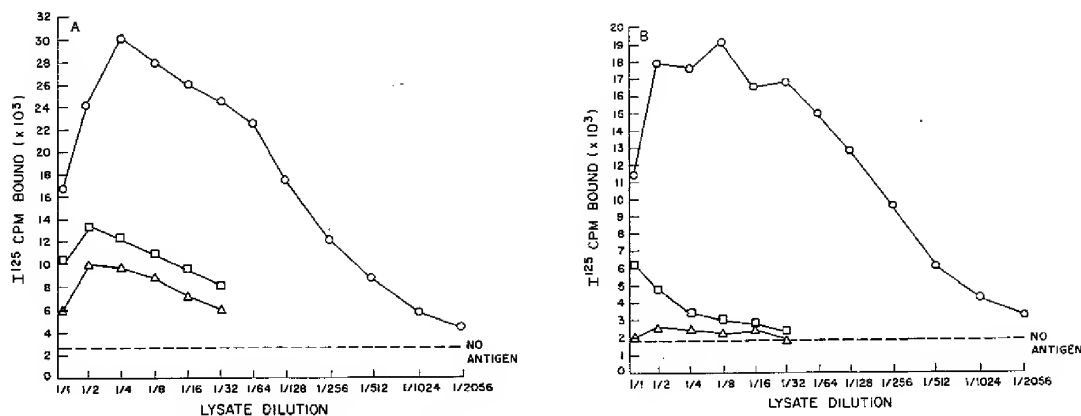
Figure 2. Binding Curve for β -Galactosidase

The wells of a polyvinyl chloride microtiter plate were coated with F(ab)₂ fragments as described in Experimental Procedures. β -Galactosidase samples (50 μ l) were added to the wells, followed by incubations with antiserum and ^{125}I -labeled protein A (*Staph. aureus*) as detailed in Experimental Procedures. The amount of β -galactosidase per well was determined by making serial dilutions from a stock (1 mg/ml) solution of purified β -galactosidase (Sigma). Samples were tested in quadruplicate; standard errors are represented by bars. (B) represents an autoradiogram of the same titration curve for β -galactosidase antigen. The microtiter plate was placed over a Kodak XR-5 film and intensifying screen for 24 hr.

some inhibition of the antibody-antigen binding reaction occurs in the presence of detergent. (The degree of inhibition was variable, depending upon how much of the detergent was transferred from the lysate into the assay well.) Inhibition by detergent is not a complete explanation for the "pro-zone" effect, however, since a small rise in the binding curve is sometimes observed in sonicated extracts as well. Reactions were run with SDS concentrations from 0.5–0.025%, indicating an optimal level of around 0.1%, since the efficiency of cell lysis was reduced at lower levels. The β -galactosidase levels in the lysates from induced lac^+ cultures are not limiting until the lysate has been substantially diluted. In Figure 3A, the undiluted lac^+ and the lac^∇ cultures can be distinguished; the binding of ^{125}I -labeled protein A molecules to the lac^∇ extract, however, is significantly above the background level.

The low level binding to protein observed in a lac^∇ extract could be due either to contaminating antibody in the rabbit antiserum directed against other bacterial antigenic determinants or to a cross-reaction of the anti- β -galactosidase with other immunologically related bacterial proteins. To distinguish between these possibilities, a lac^∇ extract was prepared by sonication and coupled to CnBr-activated Sepharose. The antiserum was passed over the lac^∇ extract affinity column, and the excluded volume fraction was tested (Figure 3B). The binding to lac^∇ lysate was significantly reduced, suggesting the presence of some contaminating antibodies in the original antiserum. All subsequent assays were performed with the absorbed β -galactosidase antiserum.

The sensitivity of the immunoassay is such that even minor contaminants are detected; they would not be expected to present a problem, however, in

Cell
684Figure 3. Comparison of lac^+ and lac^- Lysates in the Indirect Radioimmunoassay

Lysates (50 μ l) from MC1000 [lac^- (Δ)] and D7001 [lac^+] cultured in the absence (\square) or presence (\circ) of 10^{-3} M IPTG were added to the F(ab)'₂-coated wells of a microtiter plate. After a 2 hr incubation, 50 μ l of unabsorbed (A) or absorbed (B) antiserum (a 1/1000 dilution in 1% BSA) were added to the well. 2 hr later, protein A-¹²⁵I (50 μ l) was incubated in the well (100,000 cpm per well) overnight at 4°C. All samples were tested in duplicate.

identifying gene products encoded by foreign DNA. In homologous cloning systems, antiserum can be purified by absorption with extract from the appropriate deletion strain. In heterologous cloning systems, if antiserum directed against a eucaryotic protein has a cross-reaction with a bacterial protein, the cross-reactive antibody can be absorbed similarly with bacterial extract coupled to Sepharose.

The titration curve for radioactivity bound to the induced lac^+ lysate indicates that about 500 times as much immunologically reactive material is present as in the uninduced lac^+ lysate. These same lysates, assayed by the conventional colorimetric enzyme assay, show an approximately 300 fold induction ratio (Table 2). Thus the quantitation of antigen by the indirect radioimmunoassay for β -galactosidase is in general agreement with measurements of enzyme activity.

To estimate the number of cells producing gene product required for detection, lac^+ and lac^- cells were mixed in varying proportions and lysed, and the extracts were tested in the indirect radioimmunoassay (Figure 4). Based on this reconstruction experiment, the limit of detection is on the order of 1 lac^+ cell in 10^3 lac^- cells, or about 5×10^4 β -galactosidase-producing cells per well.

β -Galactosidase levels in the EK2 bacterial host, $\times 1776$, have also been examined in this assay. In L broth cultures supplemented with thymidine and diaminopimelic acid (DAP), the same SDS concentration is required to achieve efficient cell lysis.

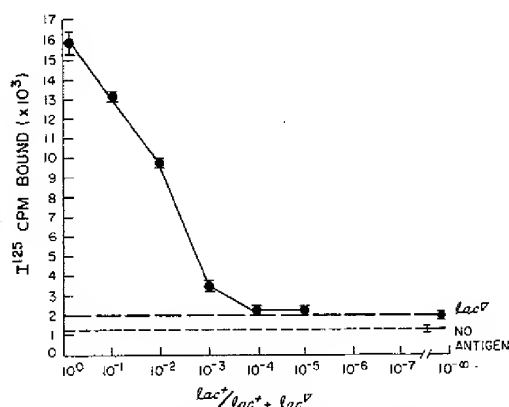
In Situ Analysis (Phage Plaques)

For the in situ analysis of β -galactosidase in bacteriophage plaques, we have adapted the immu-

Table 2. β -Galactosidase Enzyme Activity in Saturated Cultures

| Enzyme Units per ml Culture | |
|-------------------------------|---------|
| (lac^-) MC1000 | 0.98 |
| (lac^+) D7001 (Uninduced) | 14.05 |
| (lac^+) D7001 (Induced) | 3939.00 |

Saturated cultures of MC1000 (lac^-), uninduced D7001 (lac^+) and D700 (lac^+), induced by growth in the presence of 10^{-3} M IPTG, were lysed with SDS and chloroform, and assayed for β -galactosidase activity as described in Experimental Procedures.

Figure 4. Reconstruction Mixture of lac^+ and lac^- Cells

Saturated L broth cultures of MC1000 (lac^-) and D7001 (lac^+) were mixed in various proportions, and lysed with chloroform and SDS as described in Experimental Procedures. The lysates were diluted $1/2$ in 1% BSA, and 50 μ l were added to the F(ab)'₂-coated wells of a microtiter plate. After a 2 hr incubation, 50 μ l of absorbed antiserum (diluted 1/1500 in 1% BSA) were added to the well. 50 μ l of protein A-¹²⁵I (about 100,000 cpm per well) were incubated in the well overnight at 4°C, and the plate was analyzed as described in Experimental Procedures. All samples were tested in quadruplicate; standard errors are represented by bars.

noassay such that the $F(ab)'_2$ fragments are coupled to a chemically reactive cellulose filter, activated by the method of Alwine, Kemp and Stark (1977). Once the $F(ab)'_2$ fragments have been coupled to the circular cellulose filter and the reactive sites inactivated by treatment with 1 M glycine and 0.5% BSA (bovine serum albumin), the filter is applied directly to the phage plate for 1 hr. The filter is then incubated with antiserum for 3 hr and subsequently with ^{125}I -labeled protein A for 3 hr, and finally washed and overlaid with x-ray film for autoradiography. The results for a mixture of λ C1857 p5/*lac* ∇ (att)SR2 and λ C1857 plated with MC1000 (*lac* ∇) cultures on YT plates (Miller, 1972) containing the β -galactosidase stain X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside) are shown in Figure 5. All plaques, identified on the phage plate by the X-Gal stain, are clearly detected as discrete dark spots on the x-ray film. To estimate the sensitivity of the in situ plaque assays, λ phage bearing a *lac* gene, with a low efficiency promoter, λ C1857 p5(*lac* p-162), was plated alone and mixed with λ C1857 on MC1000 (Figure 6). Here, small discrete spots on the x-ray film identify λ C1857 p5 (*lac* p-162) plaques, although the β -galactosidase activity of these plaques was so low that they went undetected by the X-Gal stain in the top layer agar. The plaques, identified as λ p5(*lac* p-162) by autoradiography, were examined by picking and testing them individually in the colorimetric enzyme assay with ONPG as substrate. Since the activity in the plaques was so low, the putative λ p5(*lac* p-162)

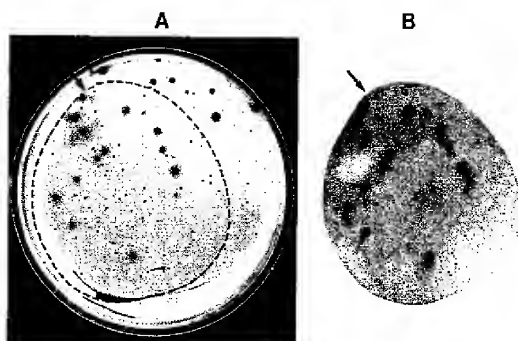


Figure 5. In Situ Radioimmunoassay for β -Galactosidase in λ C1857 p5 *lac* ∇ (att)SR2 Plaques

0.1 ml of a L broth dilution of λ C1857 p5 *lac* ∇ (att)SR2 mixed with 0.1 ml of a dilution of λ C1857 was plated with 0.1 ml of a saturated MC1000 (*lac* ∇) culture in 2 ml H top agar containing 1 mg X-Gal on YT plates. The plates were incubated at 37°C for 5 hr and then overlaid with a $F(ab)'_2$ -coated cellulose filter for 1 hr. The filter was incubated with antiserum (1/1000) for 3 hr and then with a ^{125}I -labeled protein A solution for 3 hr. The washed and dried filter was overlaid with Kodak NS-T2 X-ray film for 6 hr. (A) YT plate with X-Gal stain in top layer agar. The dotted outline indicates the area overlaid by the cellulose filter. (B) Autoradiogram of the treated filter.

plaques were also verified genetically by testing the ability of the phage to recombine with a strain carrying the small internal deletion, *s(lac z)* m15 (Langley et al., 1975) in the β -galactosidase structural gene. Recombinant (*lac* $^+$) colonies were scored on MacConkey's lactose plates. The β -galactosidase activity in λ p5(*lac* p-162) plaques measured by colorimetric enzyme assay 5 hr after plating was slightly higher than the background level of λ C1857 plaques and only about 0.05% of the activity in λ p5 *lac* plaques (Table 3).

In Situ Immunoassay for Bacterial Colonies

A mixture of a saturated MC1000 (*lac* ∇) culture and a D7001 (*lac* $^+$) culture were plated on YT plates with or without IPTG and grown up to microcolonies (about 0.5 mm in diameter) at about 1000 per

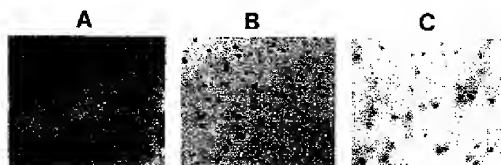


Figure 6. In Situ Radioimmunoassay for β -Galactosidase in λ C1857 p5 (*lac* p-162) Plaques

0.1 ml of an L broth dilution of λ C1857 p5 (*lac* p-162) was plated with 0.1 ml of an MC1000 (*lac* ∇) culture on YT plates containing 1 mg X-Gal in the top layer agar and overlaid with a $F(ab)'_2$ -coated cellulose filter. The filter was treated with antiserum and ^{125}I -labeled protein A, and analyzed by autoradiography as described in Figure 5. The X-ray film was exposed for 18 hr. A portion of the autoradiogram is shown in (A).

A mixture of λ C1857 p5 (*lac* p-162) and λ C1857 was plated (in a ratio of 1:1) with MC1000 (*lac* ∇) on YT plates containing X-Gal in the top layer and overlaid with a $F(ab)'_2$ -coated filter. The treated filter was analyzed by autoradiography. The X-ray film was exposed for 36 hr. A portion of the phage plate with a mixture of λ C1857 and λ C1857 p5 (*lac* p-162) plaques is shown in (B); the corresponding portion of the autoradiogram is shown in (C). The genotype of λ phage in individual plaques was verified by testing for the ability to recombine with an *E. coli* strain carrying a small internal deletion in the *lac z* gene.

Table 3. Measurement of β -Galactosidase Enzyme Activity in Bacteriophage λ Plaques

| λ Stock | Enzyme Units per Plaque | |
|---------------------------------------|-------------------------|--|
| | SE | |
| C1857 | 0.0004 \pm 0.0004 | |
| C1857 p5 <i>lac</i> ∇ (att)SR2 | 5.48 \pm 0.63 | |
| C1857 p5 (<i>lac</i> p-162) | 0.002 \pm 0.0004 | |

Individual plaques were picked 5 hr after plating and suspended in 0.5 ml "z" buffer (Miller, 1972). ONPG (50 μ l at 4 mg/ml) was added, and the reaction was terminated by the addition of 100 μ l 1 M Na_2CO_3 . Absorbance at 420 and 550 was measured. Average values were calculated from measurements of five individual plaques. Values calculated from measurements of five pooled plaques gave similar results.

plate. The colonies were lysed by exposing the plates to chloroform vapor and then overlaying them with a thin (1.5 ml) agar top layer containing lysozyme and SDS (0.025%). The F(ab)₂ fragment-coated filter was applied directly to the surface of the plate for 3 hr, treated with antiserum and ¹²⁵I-labeled protein A as described above, and analyzed by autoradiography. The results are shown in Figure 7. The induced *lac*⁺ colonies (that is, those grown on the IPTG-containing plates) are identified by large dark spots on the x-ray film. Uninduced *lac*⁺ clones, which show up as very small, discrete spots on the x-ray film, are also clearly distinguishable from the *lac*⁻ colonies. A sufficient number of cells remained viable in these microcolonies, even after treatment with chloroform vapor and the lysozyme-SDS top layer agar, so that these colonies, identified as *lac*⁺ by autoradiography, could be tested and were confirmed to be *lac*⁺ by streaks on MacConkey's agar plates. Since viable cells can be recovered from the colonies embedded in the lysozyme-SDS top layer, replica plating is not necessary for this in situ immunoassay.

Discussion

The most general method for detecting gene translation products in bacterial cells is an immunoassay, since it requires neither an intact nor a functional protein. The indirect radioimmunoassay described here for β -galactosidase as a model protein has a number of useful properties. Unlike direct radioimmunoassays, it does not require purification and labeling of specific antibody. The labeled detecting reagent, protein A, can be used in conjunction with different antisera for screening clones for a variety of gene products. The assay is simple, yet sensitive, reproducible and quantitative, and is capable of reliably detecting below 0.2

ng of β -galactosidase, or 2×10^8 molecules.

In addition, the indirect radioimmunoassay can be used for the in situ analysis of bacterial colonies and phage plaques, and by coupling the F(ab)₂ fragments to the wells of a microtiter plate, individual clones in microcultures can be assayed quantitatively or heterogeneous cultures can be analyzed by sib selection. For mass screening, the transfer of microculture lysates in a microtiter plate to the F(ab)₂-coated wells of another microtiter plate can be streamlined by using a 96 channel automatic pipetter (Cooke) and the plates monitored by autoradiography.

Uninduced *lac*⁺ cultures, estimated to contain 10–20 β -galactosidase molecules per cell (Miller, 1972) are detected easily and reliably when the lysate from 5×10^7 cells is assayed. When cultured in the presence of IPTG, 1 *lac*⁺ in 10^3 *lac*⁻ cells, or as few as 5×10^4 *lac*⁺ cells, can be identified.

In our measurements for saturated L broth cultures, induced *lac*⁺ cultures have about 300 times the enzyme activity of the uninduced lysates. Assuming, then, that induced *lac*⁺ cells contain about $3\text{--}6 \times 10^5$ molecules per cell, the indirect radioimmunoassay is capable of detecting $1.5\text{--}3.0 \times 10^8$ β -galactosidase molecules in cell lysates. This figure is in good agreement with the estimate of sensitivity based on the detection of β -galactosidase in solution in Figure 2.

The chloroform-SDS lysis method was developed to optimize the rapid mass screening of microcultures. If the number of clones is small, however, concentrated, sonicated extracts of bacterial cells can be used, thus avoiding the problem of detergent inhibition and allowing the extract from more than 10^8 cells to be measured in a single well. Under these conditions, the sensitivity of the assay should be enhanced by at least an order of magnitude, permitting the detection of clones producing



Figure 7. In Situ Radioimmunoassay for β -Galactosidase in Induced and in Uninduced Colonies of D7001 (*lac*⁺)

A mixture of MC1000 (*lac*⁻) and D7001 (*lac*⁺) cultures was plated on YT plates containing IPTG (A) or without IPTG (B) and grown up to microcolonies (about 0.5 mm in diameter) at about 1000 colonies per plate. The plates were exposed to chloroform vapors for 5 min and then overlaid with 1.5 ml of H top agar containing lysozyme (0.25 mg) and SDS (0.025%). The F(ab)₂ fragment-coated filter was applied directly to the plate surface for 2 hr and treated with antiserum and ¹²⁵I-labeled protein A as described in Figure 5. The filters were placed on Kodak NS-T2 for autoradiography. (A) represents an X-ray film exposed for 18 hr; the film shown in (B) was exposed for 24 hr.

as little as one molecule per cell or the identification of $1 \text{ } \lambda \text{ } \text{lac}^+$ in $10^4 \text{ } \lambda \text{ } \text{lac}^-$ cells.

By coupling the F(ab)'_2 fragments to a chemically activated cellulose filter, the indirect radioimmunoassay can be adapted, using autoradiography, to the in situ analysis of bacterial colonies and of phage plaques. Thus molecular cloning systems utilizing either plasmid or phage vectors can be analyzed with efficiency (as many as 10^3 clones per plate) and sensitivity. In the in situ analysis of plaques of $\lambda \text{ } \text{p5 } \text{lac}$, dark spots on the x-ray film showed a one-to-one correspondence with plaques stained by X-Gal on the plate. The size of the x-ray film spot was proportional to the size of the enzymatic stain surrounding the phage plaque.

Small plaques, which had barely detectable halos of X-Gal stain and were therefore presumably producing low levels of β -galactosidase, were easily identified as small discrete spots on the x-ray film. Thus the radioimmunoassay with autoradiographic screening is more sensitive than the X-Gal enzymatic stain for identification of β -galactosidase-containing phage plaques. The superior sensitivity of the immunoassay is also indicated by the ability to detect specific film darkening in the absence of any detectable X-Gal staining for plaques of $\lambda \text{ } \text{p5 } \text{lac } \text{p-162}$, a phage carrying a lac allele with a low efficiency promoter mutation. Plaques of $\lambda \text{ } \text{p5 } \text{lac}$ (0.2–1.0 mm in diameter) measured 5 hr after plating contained about 5.5 units of β -galactosidase activity, an average value based upon the measurement of individual plaques. In calculations based on the estimate of 300,000 units per mg protein (Miller, 1972) (1 unit = 3.6×10^9 molecules), $\lambda \text{ } \text{p5 } \text{lac}$ plaques contain about 2.0×10^{10} molecules. The β -galactosidase activity of individually tested $\lambda \text{ } \text{p5 } \text{lac } \text{p-162}$ plaques is $<0.05\%$ of the activity in $\lambda \text{ } \text{p5 } \text{lac}$ plaques. Thus the lower limit of sensitivity for the in situ λ plaque immunoassay is below 1.0×10^7 molecules.

The in situ analysis of bacterial colonies utilizes the same F(ab)'_2 -coated cellulose filter and is, in addition, dependent upon the lysis of colonies that are embedded in a top layer containing SDS and lysozyme. Large discrete spots on the x-ray film are observed corresponding to induced lac^+ colonies on plates containing IPTG. In the absence of IPTG, lac^+ colonies (uninduced) are revealed as very small, yet easily discernible spots on the x-ray film. Thus microcolonies (0.5–1.0 mm in diameter) containing only 10–20 molecules of β -galactosidase per cell are easily identified.

Unlike earlier detection methods based on immunoprecipitation, the sensitivity of the in situ indirect immunoassay is independent of the serum titer or affinity. The chemical coupling of F(ab)'_2

fragments to an insoluble matrix provides for a localized high concentration of detecting reagents and is therefore not dependent upon the diffusion of antibodies and antigen in an agar gel. Accordingly, the amount of antisera required is less than is needed for assays where detection occurs in the agar plate. In addition, since the array of colonies of plaques on a plate is only in temporary contact with the detecting antibody, the same set of clones can be screened with a panel of specific F(ab)'_2 -coated cellulose filters to monitor gene expression for a number of different proteins. This may prove useful for cloning systems in which gene families (for example, the histones) are being analyzed, and in shot-gun cloning systems where more than one gene is under investigation.

Since about 1000 individual colonies or plaques can be monitored per plate, the in situ radioimmunoassay represents a very powerful and sensitive general method for screening clones for the expression of foreign gene fragments cloned into phage or plasmid vectors. In addition, this immunoassay provides a method for examining the conditions that would allow a cloned eucaryotic DNA sequence, previously identified by nucleic acid hybridization probes, to be expressed at the protein level.

Experimental Procedures

Iodination of Protein A

Protein A (purchased from Pharmacia) was labeled with ^{125}I using chloramine T in a modification of the original method of Hunter and Greenwood (1962). 20 μl of a 1 mg/ml protein A solution in 0.5 M phosphate buffer (pH 7.0) were mixed with 20 μl of neutralized $\text{Na } ^{125}\text{I}$ (purchased from New England Nuclear) and 10 μl of a 1 mg/ml solution of chloramine T in 0.5 M phosphate buffer. After 1 min, the reaction was stopped by adding 50 μl of 1 mg/ml tyrosine in 0.5 M phosphate buffer, and 100 μl of 1% BSA were added as carrier protein. The mixture was applied immediately to a 10 ml column of Sephadex G-25 and eluted with 1% BSA. The peak fractions were pooled and stored at 4°C in the presence of 0.5% sodium azide. A fresh batch of reagent was labeled every 6–8 weeks. The specific activity of the labeled reagent, assuming no protein loss, varied between 15–30 $\mu\text{Ci}/\mu\text{g}$.

Preparation of F(ab)'_2 Fragment of Rabbit Anti- β -Galactosidase

The F(ab)'_2 fragment was derived from rabbit anti-E. coli β -galactosidase serum 484 according to a modification of the method of Madsen and Rodkey (1976). 1 ml of serum was incubated with an equal volume of saturated ammonium sulfate. The pellet resulting from two precipitation steps with 50% ammonium sulfate was resuspended in 1 ml 0.1 M sodium acetate (pH 7.0) and dialyzed overnight against the acetate buffer (pH 7.0). The protein solution was adjusted to a pH of 4.3 with 1 M HCl and incubated with 100 μl of pepsin (Worthington Biochemical Co.) at 4 mg/ml in 0.1 M acetate buffer (pH 4.3) for 8 hr at 37°C . After dialysis against PBS [0.15 M phosphate-buffered saline (pH 7.4)], the protein solution was applied to a Sephadex G-100 (Pharmacia) column and then passed over a 3 ml affinity column of protein A coupled to Cn-Br-activated Sepharose 4B to remove any residual immunoglobulin molecules bearing Fc determinants. Passage over the protein A-Sepharose column alone proved sufficient for obtaining clean preparations of F(ab)'_2 fragments.

Cell
688

Conditions for the Microtiter Plate Indirect Radioimmunoassay

The β -galactosidase immunoassay is a modification of a plate assay (Klinman et al., 1976) designed to measure serum antibody. 75 μ l of the F(ab)₂ solution (0.025 mg/ml in PBS) are incubated in the wells of a flat-bottomed, disposable, polyvinyl chloride microtiter plate (Cooke Laboratory Products) for 2 hr at room temperature. 50 μ l of antigen solution, either pure protein or cell lysate, are then incubated in the F(ab)₂-coated wells for 2 hr at room temperature. The anti- β -galactosidase serum, at a dilution of 1/500–1/2000 is added to the wells for 2 hr at room temperature. 50 μ l of a protein A ¹²⁵I solution, diluted to a final concentration of 0.1–0.3 μ g/ml and to a final activity of about 100,000–200,000 cpm per well, are incubated overnight at 4°C. Unbound reagents are removed by washing with PBS-FCS (phosphate-buffered saline with 5% fetal calf serum). All dilutions are made in 1% BSA (bovine serum albumin). ¹²⁵I-labeled protein A molecules bound to the wells are measured by cutting and counting the wells in a Searle gamma counter or by autoradiography.

Autoradiography

The microtiter plates were placed over Kodak XR-5 film and a Kodak X-Omatic Regular Intensifying Screen, and the film was developed after exposure for 20–40 hr. The treated cellulose filters were placed over Kodak NS-T2 film, and the film was developed after 6–18 hr of exposure.

Preparation and Usage of *lac*⁺ Extract-Sepharose Affinity Column

Sonicated extract prepared from 20 ml of a saturated MC1000 (*lac*⁺) culture was dialyzed overnight against coupling buffer [0.1 M Na₂CO₃, 0.3 M NaCl (pH 8.5)] and incubated overnight at 4°C with 1.3 gm (about 5 ml) of activated Cn-Br-Sepharose-4B (Pharmacia). The gel was washed with 1 M ethanolamine (pH 9.0) and then, in alternating cycles, with 0.1 M acetate buffer (pH 4.0) and 0.1 M borate buffer (pH 8.0) according to the manufacturer's instructions.

2.0 ml of 484 antiserum (diluted 1/100 in 1% BSA) were passed over a 3 ml column followed by 6 ml of 1% BSA. The absorbed serum was therefore diluted 1/400. The column was washed with 1 M propionic acid and stored in PBS in the presence of 0.1% sodium azide.

E. coli Strains

Strains MC 100 (*lac*⁺) and D7001 (*lac*⁺), obtained from Dr. Malcolm Casadaban (Stanford University), were cultured in L broth. β -Galactosidase was induced by the addition of 10^{-3} M isopropyl- β -thiogalactoside (IPTG; purchased from Sigma). X1776, obtained from Dr. Roy Curtiss III (University of Alabama), was grown in L broth, supplemented with 40 μ g/ml thymidine and 100 μ g/ml diaminopimelic acid. Cultures were lysed by the addition of chloroform (2 drops per ml) and SDS to a final concentration of 0.1% followed by vortexing for 1 min. Enzyme assays for β -galactosidase activity in the chloroform-SDS lysates were performed using ONPG (O-nitrophenyl- β -D-galactoside) as substrate according to Miller (1972).

λ Stocks

Stocks λ C1857 p5/*lac*⁺(att)SR2, λ C1857 p5/*lac*⁺p162 and λ C1857, obtained from Dr. Malcolm Casadaban, were diluted in L broth and plated with 0.1 ml of a saturated L broth culture of MC 1000 (*lac*⁺) in 2 ml of H top layer agar on YT plates (Miller, 1972). The plates were incubated at 37°C. To stain the β -galactosidase-producing plaques, 1 mg of X-Gal (5-brom-5-chloro-3-indolyl- β -D-galactoside) was added to the top layer (2 ml).

Binding F(ab)₂ Fragments to a Cellulose Filter

Whatman 540 cellulose paper was converted to ABM (aminobenzyloxy methyl) paper with treatment by 1-[(m-nitrobenzyloxy) methyl] pyridinium chloride according to the method of Alwine et al. (1977). Circles were cut from the ABM paper and converted to the reactive diazobenzyloxy methyl (DBM) cellulose paper by

treatment with HCl and NaNO₂ (Alwine et al., 1977). The activated filters were incubated with 2 ml of anti- β -galactosidase F(ab)₂ fragment (0.15 mg/ml) in Borate buffer [50 mM (pH 8)] for 18 hr at 4°C. To inactivate the reactive sites, the filters were then incubated with 5 ml of 0.1 M glycine and 0.5% BSA in PBS for 5 hr at 37°C.

Conditions for the in Situ Phage Plaque Assay

The F(ab)₂-coated cellulose filter was applied directly to the phage plate about 5 hr after pouring the top layer. After a 1 hr incubation at 37°C on the plate, the filter was removed, washed with PBS-FCS and incubated with 3 ml of anti- β -galactosidase serum (diluted 1/1000 in 1% BSA) for 3 hr. The filter was then washed, incubated with 3 ml of ¹²⁵I-labeled protein A (0.1–0.4 mg/ml; 10⁶ cpm/ml) for 3 hr, washed, dried and placed on Kodak NS-T2 x-ray film for autoradiography.

Conditions for the in Situ Bacterial Colony Assay

A mixture of MC1000 (*lac*⁺) and D7001 (*lac*⁺) cultures were plated on YT plates, with and without IPTG (2×10^{-3} M), and allowed to grow up to microcolonies (about 0.5 mm in diameter) at about 1000 colonies per plate. The colonies were lysed by exposing them to chloroform vapors for 5 min and then overlaying the plate with 1.5 ml of H top agar containing lysozyme (0.25 mg) and SDS (0.025%). The F(ab)₂ fragment-coated filter was applied directly to the surface of the top layer for 2–3 hr, removed and washed with PBS-FCS. The filter was then treated with anti- β -galactosidase serum and ¹²⁵I-labeled protein A, and analyzed by autoradiography as described above.

Acknowledgments

We are grateful to Dr. Malcolm Casadaban for his generosity in providing E. coli strains and λ phage stocks, and for many stimulating discussions. We also wish to thank Drs. George Stark, James Alwine and David Kemp for providing us with a copy of their manuscript prior to publication and with a sheet of ABM paper. We are grateful to Dr. Eli Sercarz for a gift of rabbit anti-E. coli β -galactosidase serum. Finally, we thank Ms. Carol Ostrem for expert technical assistance.

These studies were supported in part by grants from the NIH, the NSF and the American Cancer Society.

H.A.E. is the recipient of a senior postdoctoral fellowship from the American Cancer Society, California Division.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 19, 1978

References

- Alwine, J., Kemp, D. J. and Stark, G. R. (1977). Proc. Nat. Acad. Sci. USA 74, 5350–5354.
- Chang, A. C. Y., Lansman, R. A., Clayton, D. A. and Cohen, S. N. (1975). Cell 6, 231–244.
- Forsgren, A. and Sjoquist, J. (1966). J. Immunol. 97, 822–827.
- Grunstein, M. and Hogness, D. (1975). Proc. Nat. Acad. Sci. USA 72, 3961–3965.
- Hunter, W. M. and Greenwood, F. C. (1962). Nature 194, 495–496.
- Jones, K. W. and Murray, K. A. (1975). J. Mol. Biol. 96, 455–460.
- Kedes, L. H., Chang, A. C. Y., Houseman, D. and Cohen, S. N. (1975). Nature 255, 533–566.
- Klinman, N. R., Pickard, A., Sigal, N. H., Gearhart, P. J., Metcalf, E. S. and Pierce, S. K. (1976). Ann. Immunol. 127C, 489–498.
- Langle, K. E., Villarejo, M. R., Fowles, A. V., Zamenhof, B. S. and Zabin, I. (1975). Proc. Nat. Acad. Sci. USA 72, 1254–1257.

Radioimmunoassay for DNA-Translated Product Detection
689

Madsen, L. H. and Rodkey, L. S. (1976). *J. Immunol. Methods* 9, 355-361.

Meagher, R. B., Tait, R. D., Betlach, M. and Boyer, H. W. (1977). *Cell* 10, 521-536.

Morrow, J. F., Cohen, S. N., Chang, A. C. Y., Boyer, H. W., Goodman, H. M. and Holling, R. B. (1974). *Proc. Nat. Acad. Sci. USA* 71, 1743-1747.

Miller, J. H. (1972). *Experiments in Molecular Genetics* (New York: Cold Spring Harbor Laboratory).

Ratzkin, B. and Carbon, J. (1977). *Proc. Nat. Acad. Sci. USA* 74, 487-491.

Sanzey, B., Mercereau, O., Ternynk, T. and Kourilsky, P. (1976). *Proc. Nat. Acad. Sci. USA* 73, 3394-3397.

Skalka, A. and Shapiro, L. (1976). *Gene* 1, 65-79.

Struhl, K., Cameron, J. R. and Davis, R. (1976). *Proc. Nat. Acad. Sci. USA* 73, 1471-1475.

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischler, E., Rutter, W. J. and Goodman, H. M. (1977). *Science* 196, 1313-1319.

Vapnek, D., Hautala, J. A., Jacobson, J. W., Giles, N. H. and Kushner, S. R. (1977). *Proc. Nat. Acad. Sci. USA* 74, 3508-3512.

TAB JJ

Filed on behalf of Party **Goeddel**

Paper No. _____

By: Thomas E. Friebel
Michael J. Ryan
Jones Day
222 East 41st Street
New York, New York 10017-6702
Telephone: (212) 326-3939
Fax: (212) 755-7306
tefriebel@jonesday.com
mjryan@jonesday.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES

HAROU SUGANO
MASAMI MURAMATSU, and TADATSUGU TANIGUCHI

Junior Party
(Patent 5,514,567 and 5,326,859)

v.

DAVID V. GOEDEL
and ROBERTO CREA

Senior Party
(Application 07/374,311)

Patent Interference No. **105,334** (SGL)
(Technology Center 1600)

GOEDEL CLEAN COPY OF CLAIMS

GOEDDEL CLEAN COPY OF CLAIMS
INTERFERENCE No. 105,334

Pursuant to **Bd.R. 110(a)**, Party Goeddel hereby submits a clean copy of the claims pending in Goeddel's involved application Serial No. 07/374,311, as of August 25, 2006, the date the present interference was declared.

25. A DNA encoding a mature human fibroblast interferon having a total of 166 amino acids and unaccompanied by a human fibroblast interferon presequence.

26. A DNA according to claim 25, comprising the sequence:

```
ATG AGC TAC AAC TTG CTT GGA TTC CTA CAA AGA AGC AGC AAT TTT CAG
TGT CAG AAG CTC CTG TGG CAA TTG AAT GGG AGG CTT GAA TAT TGC CTC
AAG GAC AGG ATG AAC TTT GAC ATC CCT GAG GAG ATT AAG CAG CTG CAG
CAG TTC CAG AAG GAG GAC GCC GCA TTG ACC ATC TAT GAG ATG CTC CAG
AAC ATC TTT GCT ATT TTC AGA CAA GAT TCA TCT AGC ACT GGC TGG AAT
GAG ACT ATT GTT GAG AAC CTC CTG GCT AAT GTC TAT CAT CAG ATA AAC
CAT CTG AAG ACA GTC CTG GAA GAA AAA CTG GAG AAA GAA GAT TTT ACC
AGG GGA AAA CTC ATG AGC AGT CTG CAC CTG AAA AGA TAT TAT GGG AGG
ATT CTG CAT TAC CTG AAG GCC AAG GAG TAC AGT CAC TGT GCC TGG ACC
ATA GTC AGA GTG GAA ATC CTA AGG AAC TTT TAC TTC ATT AAC AGA CTT
ACA GGT TAC CTC CGA AAC.
```

27. A replicable microbial expression vehicle capable, in a transformant microorganism, of directing expression of a DNA of claim 25 or 26.

28. A microorganism transformed with a microbial expression vehicle capable, in said microorganism, of directing expression of a DNA of claim 25 or 26.

1
2 29. The microorganism of claim 28 wherein said microorganism is Escherichia coli.

3
4 30. The transformed microorganism according to claim 29 wherein said
5 microorganism is Escherichia coli K-12 strain 294.

6
7 31. A method for producing a first polypeptide having a total of 166 amino acids and
8 having the amino acid sequence of a mature human fibroblast interferon or a second polypeptide
9 having a total of 165 amino acids and having the amino acid sequence of a mature human
10 fibroblast interferon except for the amino-terminal methionine of said interferon or a mixture of
11 said first and second polypeptides, comprising culturing a microorganism transformed with a
12 microbial expression vehicle capable, in said microorganism, of directing expression of a DNA
13 of claim 25 or 26.

14
15 32. A method for producing a polypeptide having a total of 166 amino acids and
16 having the amino acid sequence of a mature human fibroblast interferon, comprising culturing a
17 microorganism transformed with a microbial expression vehicle capable, in said microorganism,
18 of directing expression of a DNA of claim 25 or 26.

19
20 33. A method for producing a polypeptide having a total of 165 amino acids and
21 having the amino acid sequence of a mature human fibroblast interferon except for the amino-
22 terminal methionine of said interferon, comprising culturing a microorganism transformed with a
23 microbial expression vehicle capable, in said microorganism, of directing expression of a DNA
24 of claim 25 or 26.

25
26 34. A DNA which consists essentially of a DNA which codes for human fibroblast
27 interferon polypeptide.

35. A DNA consisting essentially of a DNA which codes for mature human fibroblast interferon polypeptide having the amino acid sequence:

Met Ser Tyr Asn Leu Leu Gly Phe Leu Gln Arg Ser Ser Asn Phe Gln
Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg Leu Glu Tyr Cys Leu
Lys Asp Arg Met Asn Phe Asp Ile Pro Glu Glu Ile Lys Gln Leu Gln
Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile Tyr Glu Met Leu Gln
Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser Ser Thr Gly Trp Asn
Glu Thr Ile Val Glu Asn Leu Leu Ala Asn Val Tyr His Gln Ile Asn
His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu Lys Glu Asp Phe Thr
Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys Arg Tyr Tyr Gly Arg
Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser His Cys Ala Trp Thr
Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr Phe Ile Asn Arg Leu
Thr Gly Tyr Leu Arg Asn.

36. A DNA consisting essentially of a DNA according to claim 35 which has the base pair sequence:

ATG AGC TAC AAC TTG CTT GGA TTC CTA CAA AGA AGC AGC AAT TTT
TAC TCG ATG TTG AAC GAA CCT AAG GAT GTT TCT TCG TCG TTA AAA
CAG TGT CAG AAG CTC CTG TGG CAA TTG AAT GGG AGG CTT GAA TAT
GTC ACA GTC TTC GAG GAC ACC GTT AAC TTA CCC TCC GAA CTT ATA
TGC CTC AAG GAC AGG ATG AAC TTT GAC ATC CCT GAG GAG ATT AAG
ACG GAG TTC CTG TCC TAC TTG AAA CTG TAG GGA CTC CTC TAA TTC
CAG CTG CAG CAG TTC CAG AAG GAG GAC GCC GCA TTG ACC ATC TAT
GTC GAC GTC GTC AAG GTC TTC CTC CTG CGG CGT AAC TGG TAG ATA
GAG ATG CTC CAG AAC ATC TTT GCT ATT TTC AGA CAA GAT TCA TCT
CTC TAC GAG GTC TTG TAG AAA CGA TAA AAG TCT GTT CTA AGT AGA
AGC ACT GGC TGG AAT GAG ACT ATT GTT GAG AAC CTC CTG GCT AAT
TCG TGA CCG ACC TTA CTC TGA TAA CAA CTC TTG GAG GAC CGA TTA

1 GTC TAT CAT CAG ATA AAC CAT CTG AAG ACA GTC CTG GAA GAA AAA
 2 CAG ATA GTA GTC TAT TTG GTA GAC TTC TGT CAG GAC CTT CTT TTT
 3
 4 CTG GAG AAA GAA GAT TTT ACC AGG GGA AAA CTC ATG AGC AGT CTG
 5 GAC CTC TTT CTT CTA AAA TGG TCC CCT TTT GAG TAC TCG TCA GAC
 6
 7 CAC CTG AAA AGA TAT TAT GGG AGG ATT CTG CAT TAC CTG AAG GCC
 8 GTG GAC TTT TCT ATA ATA CCC TCC TAA GAC GTA ATG GAC TTC CGG
 9
 10 AAG GAG TAC AGT CAC TGT GCC TGG ACC ATA GTC AGA GTG GAA ATC
 11 TTC CTC ATG TCA GTG ACA CGG ACC TGG TAT CAG TCT CAC CTT TAG
 12
 13 CTA AGG AAC TTT TAC TTC ATT AAC AGA CTT ACA GGT TAC CTC CGA
 14 GAT TCC TTG AAA ATG AAG TAA TTG TCT GAA TGT CCA ATG GAG GCT
 15
 16 AAC
 17 TTG.

20 37. A cloned DNA consisting essentially of a DNA having the base pair sequence
 21 defined in claim 36.

23 38. A cloned DNA consisting essentially of a DNA which codes for a polypeptide
 24 having the amino acid sequence in claim 35.

26 39. A recombinant plasmid wherein a DNA which codes for the amino acid sequence:

27 Met Ser Tyr Asn Leu Leu Gly Phe Leu Gln Arg Ser Ser Asn Phe Gln
 28 Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg Leu Glu Tyr Cys Leu
 29 Lys Asp Arg Met Asn Phe Asp Ile Pro Glu Glu Ile Lys Gln Leu Gln
 30 Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile Tyr Glu Met Leu Gln
 31 Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser Ser Thr Gly Trp Asn
 32 Glu Thr Ile Val Glu Asn Leu Leu Ala Asn Val Tyr His Gln Ile Asn
 33 His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu Lys Glu Asp Phe Thr
 34 Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys Arg Tyr Tyr Gly Arg

1 Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser His Cys Ala Trp Thr
2 Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr Phe Ile Asn Arg Leu
3 Thr Gly Tyr Leu Arg Asn

4 is inserted in a vector DNA.

5

6 40. A microorganism containing the recombinant plasmid defined in claim 39.

7

8 41. The microorganism according to claim 40 which is *Escherichia coli* K-12 strain
9 294.

10

11 42. The recombinant plasmid according to claim 39 wherein said inserted DNA
12 comprises the following base pair sequence:

13 ATG AGC TAC AAC TTG CTT GGA TTC CTA CAA AGA AGC AGC AAT TTT
14 TAC TCG ATG TTG AAC GAA CCT AAG GAT GTT TCT TCG TCG TTA AAA
15

16 CAG TGT CAG AAG CTC CTG TGG CAA TTG AAT GGG AGG CTT GAA TAT
17 GTC ACA GTC TTC GAG GAC ACC GTT AAC TTA CCC TCC GAA CTT ATA
18

19 TGC CTC AAG GAC AGG ATG AAC TTT GAC ATC CCT GAG GAG ATT AAG
20 ACG GAG TTC CTG TCC TAC TTG AAA CTG TAG GGA CTC CTC TAA TTC
21

22 CAG CTG CAG CAG TTC CAG AAG GAG GAC GCC GCA TTG ACC ATC TAT
23 GTC GAC GTC GTC AAG GTC TTC CTC CTG CGG CGT AAC TGG TAG ATA
24

25 GAG ATG CTC CAG AAC ATC TTT GCT ATT TTC AGA CAA GAT TCA TCT
26 CTC TAC GAG GTC TTG TAG AAA CGA TAA AAG TCT GTT CTA AGT AGA
27

28 AGC ACT GGC TGG AAT GAG ACT ATT GTT GAG AAC CTC CTG GCT AAT
29 TCG TGA CCG ACC TTA CTC TGA TAA CAA CTC TTG GAG GAC CGA TTA
30

31 GTC TAT CAT CAG ATA AAC CAT CTG AAG ACA GTC CTG GAA GAA AAA
32 CAG ATA GTA GTC TAT TTG GTA GAC TTC TGT CAG GAC CTT CTT TTT
33

1 CTG GAG AAA GAA GAT TTT ACC AGG GGA AAA CTC ATG AGC AGT CTG
 2 GAC CTC TTT CTT CTA AAA TGG TCC CCT TTT GAG TAC TCG TCA GAC
 3
 4 CAC CTG AAA AGA TAT TAT GGG AGG ATT CTG CAT TAC CTG AAG GCC
 5 GTG GAC TTT TCT ATA ATA CCC TCC TAA GAC GTA ATG GAC TTC CGG
 6
 7 AAG GAG TAC AGT CAC TGT GCC TGG ACC ATA GTC AGA GTG GAA ATC
 8 TTC CTC ATG TCA GTG ACA CGG ACC TGG TAT CAG TCT CAC CTT TAG
 9
 10 CTA AGG AAC TTT TAC TTC ATT AAC AGA CTT ACA GGT TAC CTC CGA
 11 GAT TCC TTG AAA ATG AAG TAA TTG TCT GAA TGT CCA ATG GAG GCT
 12
 13 AAC
 14 TTG.
 15
 16

17 43. The recombinant plasmid according to claim 39 wherein said vector DNA is an
 18 *Escherichia coli* plasmid.

19

20 44. The recombinant plasmid according to claim 43 wherein said *Escherichia coli*
 21 plasmid is pBR322.

22

23 45. A process for preparing a recombinant plasmid which comprises inserting a
 24 synthesized double stranded DNA which codes for the amino acid sequence:

25 Met Ser Tyr Asn Leu Leu Gly Phe Leu Gln Arg Ser Ser Asn Phe Gln
 26 Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg Leu Glu Tyr Cys Leu
 27 Lys Asp Arg Met Asn Phe Asp Ile Pro Glu Glu Ile Lys Gln Leu Gln
 28 Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile Tyr Glu Met Leu Gln
 29 Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser Ser Thr Gly Trp Asn
 30 Glu Thr Ile Val Glu Asn Leu Leu Ala Asn Val Tyr His Gln Ile Asn
 31 His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu Lys Glu Asp Phe Thr
 32 Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys Arg Tyr Tyr Gly Arg
 33 Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser His Cys Ala Trp Thr

Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr Phe Ile Asn Arg Leu
Thr Gly Tyr Leu Arg Asn

in a vector DNA.

46. The process according to claim 45 wherein said vector DNA is an *Escherichia coli* plasmid.

47. A process according to claim 46 wherein said *Escherichia coli* plasmid is pBR322.

48. A process for producing a microorganism capable of expression of a polypeptide with interferon activity which comprises transforming a host microorganism with a replicable recombinant plasmid containing a foreign DNA which codes for the amino acid sequence:

Met Ser Tyr Asn Leu Leu Gly Phe Leu Gln Arg Ser Ser Asn Phe Gln
Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg Leu Glu Tyr Cys Leu
Lys Asp Arg Met Asn Phe Asp Ile Pro Glu Glu Ile Lys Gln Leu Gln
Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile Tyr Glu Met Leu Gln
Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser Ser Thr Gly Trp Asn
Glu Thr Ile Val Glu Asn Leu Leu Ala Asn Val Tyr His Gln Ile Asn
His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu Lys Glu Asp Phe Thr
Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys Arg Tyr Tyr Gly Arg
Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser His Cys Ala Trp Thr
Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr Phe Ile Asn Arg Leu
Thr Gly Tyr Leu Arg Asn.

49. A process according to claim 48 wherein said host microorganism is *Escherichia coli* K-12 strain 294.

50. A recombinant expression plasmid, comprising a DNA sequence encoding a mature human fibroblast interferon having a total of 166 amino acids and unaccompanied by a human fibroblast interferon presequence operably linked therein for expression by an *Escherichia coli* host, whereby expression of a mature human fibroblast interferon having a total of 166 amino acids and unaccompanied by a human fibroblast interferon presequence by said host is enabled.

51. An *Escherichia coli* host cell capable of expressing a mature human fibroblast interferon polypeptide having a total of 166 amino acids and unaccompanied by a human fibroblast interferon presequence, transformed with an expression plasmid of claim 50.

52. A method of producing a mature human fibroblast interferon polypeptide having a total of 166 amino acids and unaccompanied by a human fibroblast interferon presequence, comprising expressing the expression plasmid of claim 50 in *Escherichia coli*.

Respectfully submitted,

Dated: September 8, 2006

/Thomas E. Friebe # 29,258/

Thomas E. Friebe

Jones Day
222 East 41st Street
New York, New York 10017-6702
Tel.: (212) 326-3939
Fax: (212) 755-7306
tiefriebel@jonesday.com

TAB KK

Filed on behalf of Party **Goeddel**

Paper No. _____

By: Thomas E. Friebel
Michael J. Ryan
Jones Day
222 East 41st Street
New York, New York 10017-6702
Telephone: (212) 326-3939
Fax: (212) 755-7306
tefriebel@jonesday.com
mjryan@jonesday.com

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES

HARUO SUGANO
MASAMI MURAMATSU, and TADATSUGU TANIGUCHI

Junior Party
(Application 08/463,757)

v.

DAVID V. GOEDEL
and **ROBERTO CREA**

Senior Party
(Patent 5,460,811)

Patent Interference No. **105,337** (SGL)
(Technology Center 1600)

GOEDEL CLEAN COPY OF CLAIMS

GOEDDEL CLEAN COPY OF CLAIMS
INTERFERENCE No. 105,337

Pursuant to **Bd.R. 110(a)**, Party Goeddel hereby submits a clean copy of the claims, as issued, in Goeddel's involved Patent No. 5,460,811.

1. A composition comprising water and a nonglycosylated polypeptide having the amino acid sequence of a mature human fibroblast interferon, said nonglycosylated polypeptide having a total of 165 or 166 amino acids and said composition being free of any glycosylated human fibroblast interferon.

2. The composition of claim 1, said nonglycosylated polypeptide having the amino acid sequence

X-Ser-Tyr-Asn-Leu-Leu-Gly-Phe-Leu-Gln-Arg-Ser-Ser-Asn-Phe-Gln-Cys-Gln-Lys-Leu-Leu-Trp-Gln-Leu-Asn-Gly-Arg-Leu-Glu-Tyr-Cys-Leu-Lys-Asp-Arg-Met-Asn-Phe-Asp-Ile-Pro-Glu-Glu-Ile-Lys-Gln-Leu-Gln-Gln-Phe-Gln-Lys-Glu-Asp-Ala-Ala-Leu-Thr-Ile-Tyr-Glu-Met-Leu-Gln-Asn-Ile-Phe-Ala-Ile-Phe-Arg-Gln-Asp-Ser-Ser-Ser-Thr-Gly-Trp-Asn-Glu-Thr-Ile-Val-Glu-Asn-Leu-Leu-Ala-Asn-Val-Tyr-His-Gln-Ile-Asn-His-Leu-Lys-Thr-Val-Leu-Glu-Glu-Lys-Leu-Glu-Lys-Glu-Asp-Phe-Thr-Arg-Gly-Lys-Leu-Met-Ser-Ser-Leu-His-Leu-Lys-Arg-Tyr-Tyr-Gly-Arg-Ile-Leu-His-Tyr-Leu-Lys-Ala-Lys-Glu-Tyr-Ser-His-Cys-Ala-Trp-Thr-Ile-Val-Arg-Val-Glu-Ile-Leu-Arg-Asn-Phe-Tyr-Phe-Ile-Asn-Arg-Leu-Thr-Gly-Tyr-Leu-Arg-Asn,

1 wherein X is H or Met.

2
3 3. The composition of claim 2, said nonglycosylated polypeptide having a formula
4 molecular weight of about 20,027.

5
6 4. The composition of claim 1, 2 or 3, said composition being free of human
7 proteins.

8
9 5. The composition of claim 1, 2 or 3, said composition containing a therapeutically
10 effective amount of said nonglycosylated polypeptide and being suitable for parenteral
11 administration.

12
13 6. The composition of claim 4, said composition containing a therapeutically
14 effective amount of said nonglycosylated polypeptide and being suitable for parenteral
15 administration.

16
17 Respectfully submitted,

18 Dated: September 8, 2006

/Thomas E. Friebe # 29,258/

19 Thomas E. Friebe

20 **Jones Day**
21 222 East 41st Street
22 New York, New York 10017-6702
23 Tel.: (212) 326-3939
24 Fax: (212) 755-7306
25 tefriebe@jonesday.com

TAB LL

Gene, 10 (1980) 11–15
© Elsevier/North-Holland Biomedical Press

11

The nucleotide sequence of human fibroblast interferon cDNA

(Recombinant DNA; restriction analysis; Maxam–Gilbert technique; amino acid sequence; signal peptide)

T. Taniguchi, S. Ohno, Y. Fujii-Kuriyama and M. Muramatsu

Department of Biochemistry, Cancer Institute, Japanese Foundation for Cancer Research, Kami-Ikebukuro, Tokyo 170 (Japan)

(Received March 12th, 1980)

(Accepted March 20th, 1980)

SUMMARY

DNA synthesized by *in vitro* reverse transcription of the interferon mRNA has been cloned and amplified as recombinant DNA, TpIF319-13 (Taniguchi et al., 1979). The nucleotide sequence of this IF cDNA which consists of 770 bp (excluding the A:T tails) has been determined. The data reported predict the hitherto unknown amino acid sequence of human fibroblast interferon and its putative signal peptide.

INTRODUCTION

Interferons are species-specific glycoproteins produced by almost all vertebrates when induced by viruses as well as other inducers (Isaacs and Lindenmann, 1957; Stewart, 1979). The best characterized biological activity of interferons is the ability to elicit an antiviral state in target cells, in conjunction with the induction of *de novo* synthesis of several proteins, such as oligoadenylate synthetase, protein kinase and phosphodiesterase (Lebleu et al., 1976; Hovanessian et al., 1977; Hovanessian and Kerr, 1979; Schmidt et al., 1979). Beside their antiviral activities, interferons appear to be involved in the regulation of immune responses (Johnson, 1977; De Maeyer and De Maeyer-Guignard, 1977), and also in inhibition of tumor growth (Gresser, 1977).

At least three different species of interferons, which have been classified as Type I (acid-stable interferons induced mainly by viruses and double-stranded RNAs) and Type II (acid-labile interferon induced

efficiently by non-viral agents such as mitogens) are produced in man (Epstein, 1977; Stewart, 1979). Among Type I interferons, fibroblasts produce an interferon which is distinct immunologically as well as in target cell specificity from that produced by leukocytes, suggesting that those two interferons are coded for by distinct structural genes (Cavaliere et al., 1977; Hayes et al., 1979).

Although biological activities as well as physico-chemical properties of these interferons have been extensively studied (Knight, 1976; Rubinstein et al., 1979; Tan et al., 1979), no complete structures have so far been established, mainly due to the low amount of the proteins produced by the human cells. Very little has been learnt about the gene structure and gene organization of interferons.

In order to solve these problems, we have constructed a hybrid plasmid containing a human fibroblast interferon cDNA sequence (Taniguchi et al., 1979) and presented evidence that the cloned DNA in fact codes for human fibroblast interferon (Taniguchi et al., 1980).

In this report, we present the nucleotide sequence

Abbreviations: bp, base pairs; IF, interferon.

of the cDNA which encompasses the entire coding region as well as a part of the untranslated regions of human fibroblast interferon mRNA.

MATERIALS AND METHODS

(a) Hybrid plasmid TpIF319-13

Isolation of a bacterial clone containing the hybrid plasmid TpIF319-13 has been described in previous papers (Taniguchi et al., 1979; 1980). All cloning procedures had to be conducted in a P3 facility at the Cancer Institute, Tokyo, as required by the Guidelines for Research Involving Recombinant DNA Molecules issued in March, 1979 by the Ministry of Education, Science and Culture of Japan. (This was a totally unwarranted precaution from the scientific point of view. — Ed.)

(b) Preparation of plasmid DNA

Plasmid DNA was prepared according to the method of J. Ecsödi and C. Weissmann (personal communication), a modification of the procedure of Currier and Nester (1976).

(c) Restriction analysis

The location of endonuclease cleavage sites was determined from polyacrylamide or agarose gel electrophoresis patterns following the published procedures (Smith and Birnstiel, 1976; Tabak and Flavell, 1978).

(d) DNA sequence analysis

DNA sequence analysis was carried out by the procedure of Maxam and Gilbert (1977), with minor modifications (C. Weissmann, personal communication).

RESULTS AND DISCUSSION

Hybrid plasmid TpIF319-13, consists of a cDNA of about 800 base pairs inserted at the *EcoRI* site of pBR322. The cDNA has been shown to contain the mRNA sequence coding for the N-terminal region of

the human fibroblast interferon as well as for the signal peptide (Taniguchi et al., 1980). The locations of several restriction endonuclease cleavage sites of the cDNA were determined; a cleavage map is shown in Fig. 1. After digestions of either intact plasmid or isolated fragments containing the cDNA insert by one or more restriction enzymes, the resulting DNA fragments were labeled at their 5' end with [32 P]-phosphate, and the nucleotide sequence analysis was carried out according to the procedure of Maxam and Gilbert (1977). The sequencing strategy is presented in Fig. 1. The whole cDNA sequence, encompassing the entire coding region of the interferon as well as a part of the non-translated region, was determined and is presented in Fig. 2.

As has been reported in a previous paper (Taniguchi et al., 1980), the 5' terminal region of the interferon mRNA, as deduced from the cDNA, contains at least 6 nucleotides before the first AUG codon in phase. If this AUG functions as an initiator, the primary translation product of human fibroblast interferon mRNA contains a signal peptide consisting of 21 amino acids which would eventually be cleaved off during secretion. Alternatively, there may be another AUG sequence farther upstream in a segment of the mRNA which was not copied into the TpIF319-13 cDNA. As shown previously, the mRNA

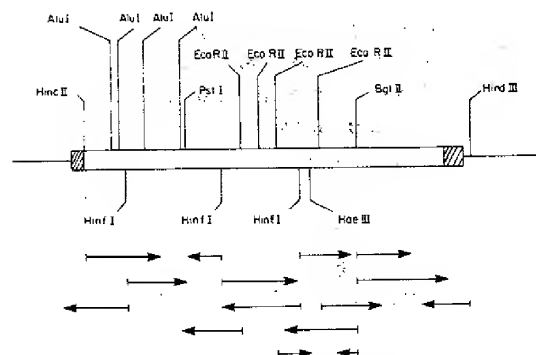


Fig. 1. Restriction map of human fibroblast interferon cDNA insert and adjacent regions of TpIF319-13. The location of endonuclease cleavage sites were determined from gel electrophoresis patterns following simultaneous and sequential digestions of either intact plasmid or isolated fragments containing cDNA insert. The rectangle represents the cDNA insert, and the shaded area the A-T junctions. The strategy of the DNA sequence analysis is presented in the lower part of the figure. Arrows indicate the direction and extent of sequencing of each fragment analysed; the vertical bar at the end of each arrow represents the 5'-terminal, labeled phosphate.

as for the
e locations
ge sites of
p is shown
plasmid or
ert by one
ting DNA
ith [³²P]-
analysis was
Maxam and
presented
compassing
as well as
determined

er (Tanigu-
f the inter-
A, contains
UG codon
itiator, the
fibroblast
peptide con-
eventually be
there may
tream in a
ied into the
the mRNA

erferon cDNA
he location of
om gel electro-
nd sequential
ted fragments
ents the cDNA
s. The strategy
the lower part
and extent of
tical bar at the
labeled phos-

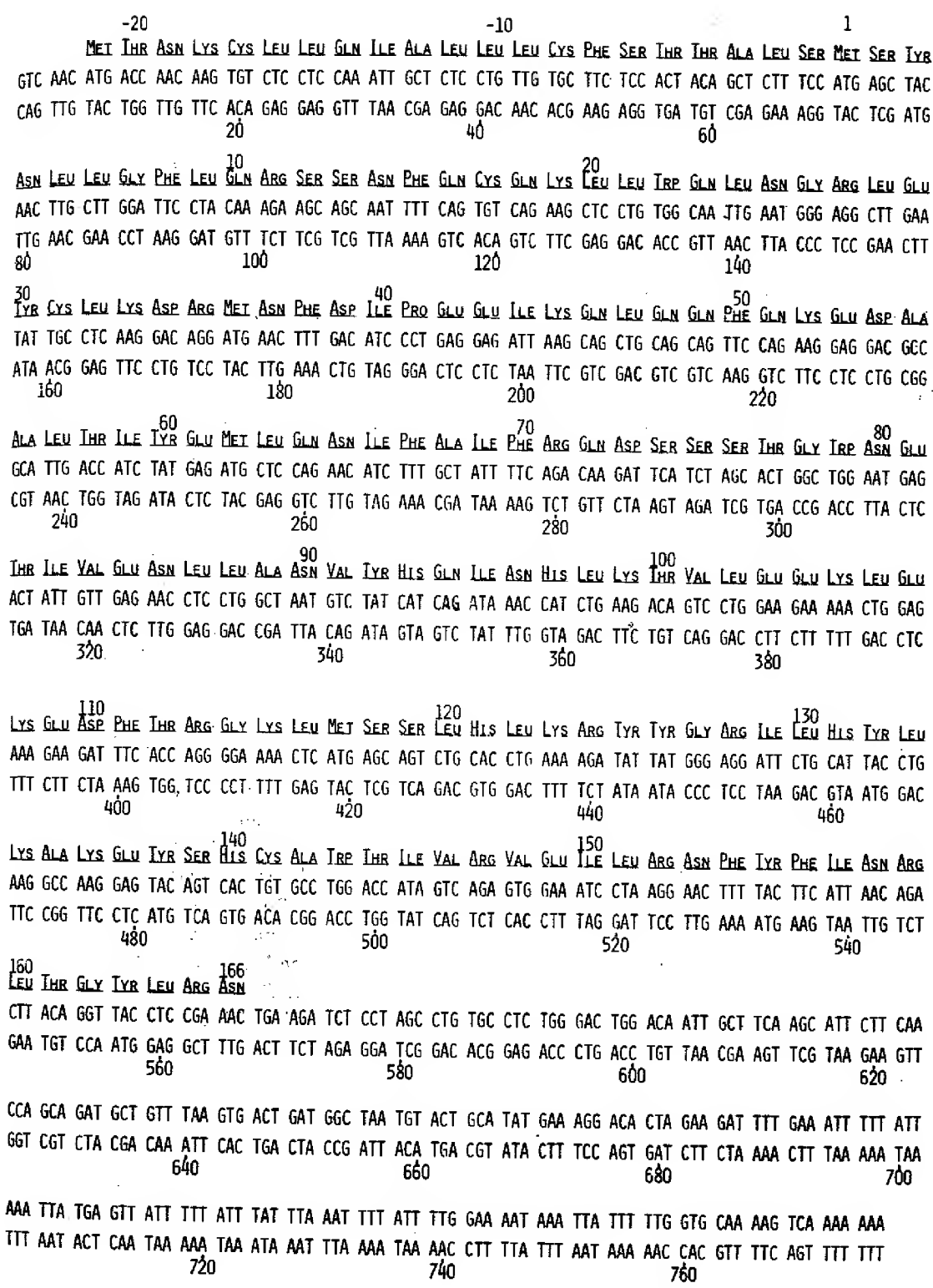


Fig. 2. Nucleotide sequence of the human fibroblast interferon cDNA and the predicted amino acid sequence of the pre-interferon molecule. Determination of the DNA sequence was carried out according to the procedure of Maxam and Gilbert (1977). (cf. Fig. 1). The DNA sequence permits prediction of the entire amino acid sequence for human fibroblast interferon (amino acids 1-166) and its putative signal peptide (amino acids -21 to -1) as shown in the line above the DNA sequences.

sequence deduced from the cDNA predicts the synthesis of a polypeptide whose N-terminal amino acid sequence is identical to that of human fibroblast interferon (Knight et al., 1980; Taniguchi et al., 1980). The sequence coding for mature interferon thus begins with a methionine codon which is followed by codons for 165 amino acids and terminates by UGA. The complete amino acid sequence of human fibroblast interferon predicted from the cDNA sequence is also presented in Fig. 2.

The amino acid composition of the interferon reported by Knight et al., (1980) agrees largely with that deduced from the cDNA sequence. The molecular weight of the human fibroblast interferon polypeptide was calculated to be 20040; this is to be compared with the values of 20000 and 26000 for glycosylated fibroblast interferon (Knight, 1976; Hayes et al., 1979) and 16000 for unglycosylated interferon (Havell et al., 1977), estimated from electrophoretic mobilities in sodium dodecylsulfate-polyacrylamide gels. The protein structure deduced from the DNA sequence agrees with the physicochemical properties of human fibroblast interferon in the following respects. Hydrophobic amino acids such as leucine (25), isoleucine (11), phenylalanine (9) and valine (5) are abundant and this is consistent with the well-known hydrophobicity of the interferon (Knight, 1976; Tan et al., 1979). There are three cysteine residues (positions 17, 31 and 141) in human fibroblast interferon which is consistent with the notion that S-S bonds may play a role in the maintenance of the interferon activity (Tan et al., 1979).

The sequence analysis of the coding region of human fibroblast interferon mRNA indicates that codon utilisation is non-random, with certain preferred codon choices for some of the amino acids (Fig. 3). The use of certain codons such as CUG and CUC for leucine, and AAG for lysine seems to be favored in many eukaryotic mRNAs examined to date (Seeburg et al., 1977; Grantham et al., 1980). It is also remarkable that codons containing the sequence CG occur quite infrequently: no CGX or XCG codon occurs in the interferon mRNA except one CGA for arginine. In addition, of the 49 codons ending with C, only 2 are followed by a codon beginning with G. This observation seems to apply to all eukaryotic mRNAs (Russell et al., 1976; Fiers et al., 1978). It may be worth noting that the AUA codon for isoleucine is not frequently used in other mRNAs

| | | | | | | | | | | | |
|-----|-----|-------|-----|-----|------|-----|-----|------|-----|-----|------|
| PHE | UUU | 4 | SER | UCU | 1 | TYR | UAU | 5 | CYS | UGU | 2(1) |
| | UUC | 5(1) | | UCC | 0(2) | | UAC | 5 | | UGC | 1(1) |
| LEU | UUA | 0 | | UCA | 1 | OC | UAA | 0 | OP | UGA | 1 |
| | UUG | 3(1) | | UCG | 0 | AM | UAG | 0 | TRP | UGG | 3 |
| LEU | CUU | 3(1) | PRO | CCU | 1 | HIS | CAU | 3 | ARG | CGU | 0 |
| | CUC | 7(3) | | CCC | 0 | | CAC | 2 | | CGC | 0 |
| | CUA | 2 | | CCA | 0 | GLN | CAA | 3(1) | | CGA | 1 |
| | CUG | 10(1) | | CCG | 0 | | CAG | 8 | | CGG | 0 |
| ILE | AUU | 5(1) | THR | ACU | 2(1) | ASN | AAU | 4 | SER | AGU | 2 |
| | AUC | 4 | | ACC | 3(1) | | AAC | 8(1) | | AGC | 5 |
| | AUA | 2 | | ACA | 1(1) | LYS | AAA | 4 | ARG | AGA | 5 |
| MET | AUG | 4(1) | | ACG | 0 | | AAG | 7(1) | | AGG | 5 |
| VAL | GUU | 1 | ALA | GCU | 2(2) | ASP | GAU | 2 | GLY | GGU | 1 |
| | GUC | 3 | | GCC | 3 | | GAC | 3 | | GGC | 1 |
| | GUA | 0 | | GCA | 1 | GLU | GAA | 5 | | GGA | 2 |
| | GUG | 1 | | GCG | 0 | | GAG | 8 | | GGG | 2 |

Fig. 3. Codon utilisation in human fibroblast interferon mRNA. The numbers indicate how many times each codon is used in the mRNA segment coding for the mature interferon polypeptide. The numbers in parentheses indicate the frequency of the codons used in the putative signal peptide.

(Grantham et al., 1980), whereas 2 out of 11 isoleucines are coded for by AUA in the case of human fibroblast interferon.

At the 3'-end of the mRNA there is a long A + T-rich untranslated segment comprising the sequence AATAAAA, which appears to be common to all polyadenylated mRNAs of eukaryotes studied up to date (Proudfoot and Brownlee, 1976).

One interesting aspect regarding human interferons is that the fibroblast and leukocyte interferons exhibit a number of different properties (Gresser et al., 1974; Havell et al., 1975; Berg et al., 1975; Vilček et al., 1977). Since both cDNA clones are available (Taniguchi et al., 1979; Nagata et al., 1980) and have been sequenced (see also preceding paper), it became possible to compare the structures of the two interferon polypeptides and to determine that human fibroblast and leukocyte interferons are structurally related (to be published).

ACKNOWLEDGEMENTS

We thank Dr. H. Sugano for his continuous support and interest. We are also indebted to Drs. C. Weiss-

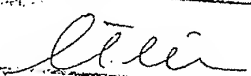
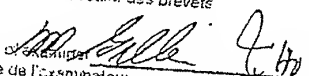
mann and A. Fuse for valuable discussions and advice. Thanks are also due to Ms. Y. Taniguchi for typing the manuscript.

REFERENCES

- Berg, K., Ogburn, C.A., Paucker, K., Mogensen, K.E. and Cantell, K.: Affinity chromatography of human leukocyte and diploid cell interferons on sepharose-bound antibodies. *J. Immunol.* 114 (1975) 640-644.
- Cavalieri, R.L., Havell, E.A., Vilček, J. and Pestka, S.: Synthesis of human interferon by *Xenopus laevis* oocytes: two structural genes for interferons in human cells. *Proc. Natl. Acad. Sci. USA* 74 (1977) 3287-3291.
- Currier, T.C. and Nester, E.W.: Isolation of covalently closed circular DNA of high molecular weight from bacteria. *Analyt. Biochem.* 76 (1976) 431-441.
- De Maeyer, E. and De Maeyer-Guignard, J.: Effect of interferon on cell-mediated immunity. *Texas Repts. Biol. Med.* 35 (1977) 370-374.
- Epstein, L.B.: Mitogen and antigen induction of interferon in vitro and in vivo. *Texas Repts. Biol. Med.* 35 (1977) 42-56.
- Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. and Ysebaert, M.: Complete nucleotide sequence of SV40 DNA. *Nature* 273 (1978) 113-120.
- Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pavé, A.: Codon catalog usage and the genome hypothesis. *Nucl. Acids Res.* 8 (1980) r49-r62.
- Gresser, I., Bandu, M.T., Brouty-Boye, D. and Tovey, M.: Pronounced antiviral activity of human interferon on bovine and porcine cells. *Nature* 251 (1974) 543-545.
- Gresser, I.: On the mechanism of the antitumor effects of interferon. *Texas Repts. Biol. Med.* 35 (1977) 394-405.
- Havell, E.A., Berman, B., Ogburn, C.A., Berg, K., Paucker, K. and Vilček, J.: Two antigenically distinct species of human interferon. *Proc. Natl. Acad. Sci. USA* 72 (1975) 2185-2190.
- Havell, E.A., Yamazaki, S. and Vilček, J.: Altered molecular species of human interferon produced in the presence of inhibitors of glycosylation. *J. Biol. Chem.* 252 (1977) 4425-4433.
- Hayes, T.G., Yip, Y.K. and Vilček, J.: Le interferon production by human fibroblasts. *Virology*, 98 (1979) 351-363.
- Hovanessian, A.G., Brown, R.E., Kerr, I.: Synthesis of low molecular weight inhibitor of protein synthesis with enzyme from interferon-treated cells. *Nature* 268 (1977) 537-539.
- Hovanessian, A.G. and Kerr, I.: The (2'-5')oligoadenylate (pppA2'5'A2'-5'A) synthetase and protein kinase(s) from interferon-treated cells. *Eur. J. Biochem.* 93 (1979) 515-526.
- Isaacs, A. and Lindenmann, J.: Virus interference, I. The interferon. *Proc. Roy. Soc. B* 147 (1957) 258-267.
- Johnson, H.M.: Effect of interferon on antibody formation. *Texas Repts. Biol. Med.* 35 (1977) 375-377.
- Knight, Jr., E.: Interferon purification and initial characterization from human diploid cells. *Proc. Natl. Acad. Sci. USA* 73 (1976) 520-523.
- Knight, Jr., E., Hunkapiller, M.W., Korant, B.D., Hardy, R.W.F. and Hood, L.E.: Human fibroblast interferon: Amino acid analysis and amino terminal amino acid sequence. *Science* 207 (1980) 525-526.
- Lebleu, B., Sen, G.C., Shaila, S., Cabrer, B. and Lengyel, P.: Interferon, double-stranded RNA, and protein phosphorylation. *Proc. Natl. Acad. Sci. USA* 73 (1976) 3107-3111.
- Maxam, A.M. and Gilbert, W.: A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* 74 (1977) 560-564.
- Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsödi, J., Boll, W., Cantell, K. and Weissmann, C.: Synthesis in *E. coli* of a polypeptide with human leukocyte interferon activity. *Nature* 284 (1980) 316-320.
- Proudfoot, N.J. and Brownlee, G.G.: 3'-Non-coding region sequences in eukaryotic messenger RNA. *Nature* 263 (1976) 211-214.
- Rubinstein, M., Rubinstein, S., Familletti, P.C., Miller, R.S., Waldman, A.A. and Pestka, S.: Human leukocyte interferon: Production, purification to homogeneity, and initial characterization. *Proc. Natl. Acad. Sci. USA* 76 (1979) 640-644.
- Russell, G.J., Walker, P.M.B., Elton, R.A. and Subak-Sharpe, J.H.: Doublet frequency analysis of fractionated vertebrate nuclear DNA. *J. Mol. Biol.* 108 (1976) 1-23.
- Seeburg, P.H., Shine, J., Martial, J.A., Baxter, J.D. and Goodman, H.M.: Nucleotide sequence and amplification in bacteria of structural gene for rat growth hormone. *Nature* 270 (1977) 486-494.
- Schmidt, A., Chernajovsky, Y., Shulman, L., Federman, P., Berissi, H. and Revel, M.: An interferon-induced phosphodiesterase degrading (2'-5') oligoadenylate and C-C-A terminus of tRNA. *Proc. Natl. Acad. Sci. USA* 76 (1979) 4788-4792.
- Smith, H.O. and Birnstiel, M.L.: A simple method for DNA restriction site mapping. *Nucl. Acids Res.* 3 (1976) 2387-2399.
- Stewart, II, W.E.: *The Interferon System*, Springer-Verlag, Vienna, 1979.
- Tabak, H.F. and Flavell, R.A.: A method for the recovery of DNA from agarose gels. *Nucl. Acids Res.* 5 (1978) 2321-2332.
- Tan, Y.H., Barakat, F., Berthold, W., Smith-Johannsen, H. and Tan, C.: The isolation and amino acid/sugar composition of human fibroblast interferon. *J. Biol. Chem.* 254 (1979) 8067-8073.
- Taniguchi, T., Sakai, M., Fujii-Kuriyama, Y., Muramatsu, M., Kobayashi, S. and Sudo, T.: Construction and identification of a bacterial plasmid containing the human fibroblast interferon gene sequence. *Proc. Japan Acad.* 55, Ser. B (1979) 464-469.
- Taniguchi, T., Fujii-Kuriyama, Y. and Muramatsu, M.: Molecular cloning of human fibroblast interferon cDNA. *Proc. Natl. Acad. Sci. USA* (1980) in press.
- Vilček, J., Havell, E.A. and Yamazaki, S.: Antigenic, physicochemical and biologic characterization of human interferon. *Proc. Natl. Acad. Sci. USA* 74 (1977) 703-710.

TAB MM

Mar - Apr. 1980

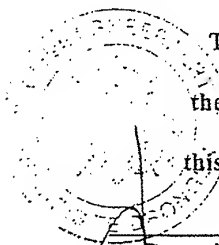
| | | |
|---|--|------|
| Opened | Dec 30 | 2002 |
|  | | |
| Commissioner of Patents Commissaire des brevets | | |
| In presence of Examiner en présence de l'examineur |  | |

Human leukocyte and fibroblast interferons are structurally related.

Tadatsugu Taniguchi⁺, Ned Mantel^o, Marco Schwarzstein^o,
Shigekazu Nagata^o, Masami Muramatsu⁺ and Charles Weissmann^o

^oInstitut für Molekularbiologie I, Universität Zürich,
8093 Zürich, Switzerland

⁺Dept. of Biochemistry, Cancer Institute, Japanese Foundation
for Cancer Research, Tokyo 170, Japan



This is EXHIBIT FIERS-40
to
the Affidavit of Walter C. Fiers
sworn before me
this 19th day of November, 2001

Sugano EXHIBIT 2064
Sugano v. Goeddel
Interference No. 105,334 and 105,337

SUMMARY

The coding sequences of the cDNAs of cloned human leukocyte interferon I and human fibroblast interferon show homologies of 45% at the nucleotide and 29% at the amino acid level. We estimate that the two genes were derived from a common ancestor about 500 to 1,000 million years ago.

The acid-stable human interferons are subdivided into two major groups, namely fibroblast interferons (F-IF) and leukocyte interferons (Le-IF); these are the major components of the interferons produced by induced fibroblasts and leukocytes, respectively. Some cells, such as the lymphoblastoid Namalva cell line, produce a mixture of 90% Le-IF and 10% F-IF (1, 2). The two interferon types have several features in common: both are glycoproteins with molecular weights ranging from 16'000 to 26'000 (3-9), the induction and shut-off of their synthesis appears to be under similar control (6), and at least some of the responses elicited in target cells are similar, such as induction of an antiviral state, which is accompanied by increased synthesis of several proteins (10-13). Nonetheless, the two kinds of interferons differ in many respects. Antibodies directed against Le-IF do not neutralize F-IF and vice-versa (14), the target cell specificities of the two IFs differ (15), and the sequences of the 13 amino terminal amino acids of F-IF and of Le-IF (from lymphoblastoid cells) show no homology (16, 22). Although Le-IF and F-IF are encoded by different mRNA species (17), it is not known whether these mRNAs are transcribed from distinct genes or whether they arise from the same gene via a common precursor which is processed or spliced in different modes.

We have recently cloned and sequenced one species each of Le-IF (Le-IF I) (18,19) and F-IF cDNA (20,21). A second

species of Le-IF (Le-IF II) cDNA has recently been identified (M. Streuli, S. Nagata and C. Weissmann, unpublished results).

In Fig. 1 the nucleotide sequences of Le-IF I and F-IF cDNA were aligned so that the AUGs closest to their 5' termini coincided. From the amino terminal sequence published for F-IF (16) and lymphoblastoid Le-IF (22) one can determine that in the case of F-IF the 21st codon following the initiation triplet and in the case of Le-IF the 23rd codon represents the first amino acid of the interferon polypeptide. Presumably the stretch in between encodes a signal peptide. Since the putative signal peptide of Le-IF comprises 23 and that of F-IF 21 amino acids, the IF polypeptides, as aligned in Fig. 1, are shifted by two residues relative to their termini. In this alignment, 48 of 166 positions (29%) of the interferon polypeptides have identical amino acids. To plot the degree of homology between the F-IF and Le-IF as function of the map distance, the sequence was subdivided into segments of 8 amino acids (or 24 nucleotides), each overlapping by 4 amino acids (or 12 nucleotides) with the neighboring segments, and the percent coincidence of amino acids (and nucleotides) for each segment was determined (cf. van Ooyen et al., ref. 23). As seen in Fig. 2, amino acid sequences show three domains of homology. The first one, with the least degree of homology, corresponds to the putative signal sequence, which is rich in hydrophobic residues and has 4 identical amino acid positions out of 21; the second domain, between the 28th and 80th amino acid (counted on the Le-IF sequence), has 21 identical residues out of 51 (41% homology) and the third, between

positions 115 and 151 (Le-IF sequence), has 19 out of 35 identical residues (54%). The longest stretches of contiguous conserved amino acids are Gln-Phe-Gln-Lys (pos. 47-50 of Le-IF and 49-52 of F-IF) and Cys-Ala-Trp (pos. 139-141 and pos. 141-143, respectively). The latter sequence is notable because it comprises Cys and Trp, which are preferentially conserved in related proteins (24). Table 1 shows that conservation was highest between the interferon polypeptides (not considering the signal sequences) for Trp, Phe, Arg, Cys and Tyr residues, in agreement with the general experience that the amino acids most likely to be conserved between related proteins are Trp > Cys > Tyr > Arg > Phe, His (24). Even where amino acids are conserved, the codons show one or more nucleotide changes in half the instances. The codons of 3 out of 7 conserved Leu residues are non-related, as are 2 of 4 codons pertaining to conserved Ser residues. This suggests that there is a strong selective pressure favoring the conservation of several amino acids. It is quite likely that at least some of the conserved amino acids are essential for a function common to Le-IF and F-IF, perhaps the induction of the virus-resistant state in the target cell. These findings may provide guidelines for the tailoring of modified (25), possibly shorter polypeptides possessing certain activities of interferon.

The nucleic acid sequences show an average homology of 43% in the domain of the signal sequence and of 45% in the interferon polypeptide sequence. On a random basis, about 25% of the nucleotide positions should

coincide. Within the interferon coding sequence, the nucleotide homologies are more evenly distributed than the amino acid homologies. However, one may distinguish, albeit to a less pronounced degree, the same two blocks of similarity noted for the amino acids. The longest region without mismatches extends for 13 nucleotides (cf. 47th to 51st codon of Le-IF vs. 49th to 53rd codon of F-IF). There are, in addition, sequences of 17, 18 and 20 nucleotides with 3, 3 and 4 mismatches, respectively. The heteropolymeric 3'-terminal non-coding region of Le-IF cDNA has 242 nucleotides, and is longer by 39 residues than its counterpart in F-IF cDNA. In aligning the two sequences four gaps were introduced to maximize homology, as described by van Ooyen et al.⁽²³⁾ Thereby, several segments were matched with 29 to 41% homology. The introduction of gaps in the alignment may be justified in view of the arguments presented previously, that introns and non-coding regions of reduplicated genes diverge as a consequence of block insertions and/or deletions in the course of evolution (23,26).

It is unlikely that the extent of homology between Le and F-IF cDNA would allow meaningful crosshybridization between the two species.

On the basis of our findings there is no doubt that Le-IF and F-IF genes are derived from a common ancestral sequence. When did the separation of these genes occur? Human α and β globin show 57% amino acid mismatches, and human β -globin and myoglobin, as well as α -globin and myoglobin, 76% mismatches.

If the rate of divergence of interferons and globins is comparable (however, cf. p. 50, ref. 24, for proteins showing both higher and lower rates) then the separation of interferon genes occurred after that of myoglobin and hemoglobins and before that of α - and β -globins, i.e. between 500 and 1000 million years ago (24). The interferon genes may thus be about as old as the vertebrates (27).

ACKNOWLEDGEMENTS

T.T. and M.M. are indebted to Dr. H. Sugano for his continuous support and interest. The work at the University of Zürich was supported by Biogen S.A. and the Schweizerische Nationalfonds.

REFERENCES

1. Havell, E.A., Yip, Y.K. & Vilček, J. J. gen. Virol. 38, 51-59 (1977).
2. Paucker, K. The Interferon System, Texas Reports on Biology and Medicine (Baron, S. & Dianzani, F., eds.) 35, 23-28 (1977).
3. Bridgen, P.J., Anfinson, C.B., Corley, L., Bose, S., Zoon, K.C., Ruegg, U.T. & Buckler, C.E. J. Biol. Chem. 252, 6585-6587 (1977).
4. Knight, E., jr. Proc. Natl. Acad. Sci. USA 73, 520-523 (1976).
5. Tan et al., J. Biol. Chem. 254, 8067-8073 (1979).
6. Hayes, T.G., Yip, Y.K. & Vilček, J. Virology 98, 351-363 (1979).
7. Rubinstein, M., Rubinstein, S., Familletti, P.C., Miller, R.S., Waldman, A.A. & Pestka, S. Proc. Natl. Acad. Sci. USA 76, 640-644 (1979).
8. Stewart, W.E., II, & Wiranowska-Stewart, M. Virology 97, 473-476 (1979).
9. Havell, E.A., Yamazaki, S. & Vilček, J. J. Biol. Chem. 252, 4425-4427 (1977).
10. Lebleu, B., Sen, G.C., Shaila, S., Cabrer, B. & Lengyel, P. Proc. Natl. Acad. Sci. USA 73, 3107-3111 (1976).
11. Hovanessian, A.G., Brown, R.E. & Kerr, I.M. Nature 268, 537-540 (1977).
12. Hovanessian, A.G. & Kerr, I.M. Eur. J. Biochem. 93, 515-526 (1979).
13. Schmidt, A., Chernajovsky, Y., Shulman, L., Federman, P., Berissi, H. & Revel, M. Proc. Natl. Acad. Sci. USA 76, 4788-4792 (1979).

14. Havell, E.A., Berman, B., Ogburn, C.A., Berg, K., Paucker, K. & Vilček, J. Proc. Natl. Acad. Sci. USA 72, 2185-2187 (1975).
15. Stewart, W.E., II, In "The Interferon System" (Springer-Verlag, Wien & New York), pp. 134-145 (1979).
16. Knight, E., jr., Hunkapiller, M.W., Korant, B.D., Hardy, R.W.F. & Hood, L.E. Science 207, 525-526 (1980).
17. Cavalieri, R.L., Havell, E.A., Vilček, J. & Pestka, S. Proc. Natl. Acad. Sci. USA 74, 3287-3291 (1977).
18. Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsödi, J., Boll, W., Cantell, K. & Weissmann, C. Nature, in press (1980).
19. Mantei, N., Schwarzstein, M., Streuli, M., Panem, S., Nagata, S. & Weissmann, C., Gene, in press (1980).
20. Taniguchi, T., Sakai, M., Fujii-Kuriyama, Y., Muramatsu, M., Kobayashi, S. & Sudo, T. Proc. Jap. Acad. Ser. B, 55, 464-469 (1979).
21. Taniguchi, T., Ohno, S., Fujii-Kuriyama, Y. & Muramatsu, M. Gene, in press (1980).
22. Zoon, K.C., Smith, M.E., Bridgen, P.J., Anfinsen, C.B., Hunkapiller, M.W. & Hood, L.E. Science 207, 527-528 (1980).
23. van Ooyen, A., van den Berg, J., Mantei, N. & Weissmann, C. Science 206, 337-344 (1979).
24. Dayhoff, M.O., In "Atlas of Protein Sequence and Structure" (Natl. Biomedical Res. Found., Washington, D.C.) Vol. V (1972).
25. Müller, W., Weber, H., Meyer, F. & Weissmann, C. J. Mol. Biol 124, 343-358 (1978).

26. Konkel, D.A., Maizel, J.V., jr. & Leder, P. Cell 18, 865-873 (1979).
27. Kurtén, B. In "Ideas in modern biology" Proceedings vol. 6, XVI International Congress of Zoology (Moore, J.A., ed.), Natural History Press, N.Y. (1965) pp. 327-354.

TABLE 1 Conservation of amino acids in leukocyte and fibroblast interferon.*

| | F-IF | Le-IF | Conserved amino acids | Number of changes in codon of conserved amino acids | | | |
|-----|------------|------------|--------------------------|--|-----------|----------|----------|
| | | | | 0 | 1 | 2 | 3 |
| Leu | 25 | 22 | 8 | 1 | 4 | 3 | |
| Cys | 3 | 5 | 2 | 1 | 1 | | |
| AsN | 12 | 6 | 1 | 1 | | | |
| Arg | 11 | 12 | 5 | 1 | 3 | 1 | |
| Phe | 9 | 8 | 4 | 2 | 2 | | |
| Pro | 1 | 6 | 1 | | 1 | | |
| GLN | 11 | 10 | 3 | 3 | | | |
| Lys | 11 | 8 | 3 | 2 | 1 | | |
| Ala | 6 | 10 | 2 | 2 | | | |
| Glu | 13 | 15 | 4 | 4 | | | |
| Ile | 11 | 7 | 3 | 2 | 1 | | |
| Ser | 9 | 13 | 4 | | 2 | 1 | 1 |
| Trp | 3 | 2 | 2 | 2 | | | |
| Tyr | 10 | 4 | 4 | 1 | 3 | | |
| Val | 5 | 6 | 1 | 1 | | | |
| Asp | 5 | 11 | 1 | 1 | | | |
| Thr | 6 | 9 | 0 | | | | |
| Gly | 6 | 3 | 0 | | | | |
| Met | 4 | 6 | 0 | | | | |
| His | 5 | 3 | 0 | | | | |
| | <u>166</u> | <u>166</u> | <u>48</u> | <u>24</u> | <u>18</u> | <u>5</u> | <u>1</u> |

*The data are from Taniguchi et al. (ref. 21) and Mantei et al. (ref. 19).

FIGURE LEGENDS

Fig. 1 Comparison of the nucleotide sequences of human leukocyte interferon I (Le-IF I) and human fibroblast interferon cDNA and of the derived amino acid sequences. The sequences are from Mantei et al. (19) and Taniguchi et al. (21). They were aligned to give maximal homology. Identical amino acids are framed, identical nucleotides are marked by a dot. S1 to S23 indicate the amino acids of the putative signal sequence; 1 to 166 the amino acids of the interferon polypeptides.

Fig. 2 Similarity of the nucleotide and amino acid sequences of human leukocyte interferon I and fibroblast interferon. The sequences shown in Fig. 1 were subdivided in segments of 8 amino acids or 24 nucleotides, each overlapping by 4 and 12 residues, respectively, with the neighboring segments. The percentage of coincident residues was plotted as a function of map position. Open vertical blocks, nucleotides; filled vertical blocks, amino acids. L-IF, leukocyte interferon cDNA; F-IF, fibroblast interferon cDNA; lines, non-coding sequences; hatched bars, putative signal peptide; open bars, interferon polypeptide.

Fig. 1

23
 G CT CTA GGT TCA GAG TCA CCC ATC TCA GCA AGC CCA GAA GTA TCT GCA ATA TCT ACG ATG GCC TCG CCC TTT
 S1
 MET ALA SER PRO PHE
 MET THR ASN LYS CYS
 GTC AAC ATG ACC AAC AAG TGT

S10 S20
 1
 GCT TTA CTG ATG GTC CTG GTG GTG CTG AGC TGC AAG TCA AGC TGC TCT CTG GGC TGT GAT CTC CCT GAG ACC
 ALA LEU LEU MET VAL LEU VAL VAL LEU SER CYS LYS SER SER CYS SER LEU GLY CYS ASP LEU PRO GLU THR
 LEU LEU GLN ILE ALA LEU LEU LEU CYS PHE SER THR THR ALA LEU MET SER TYR ASN LEU LEU GLY PHE
 CTC CTC CAA ATT GCT CTC CTC CTG TTG TGC TTC TCC ACT ACA GCT CTT TCC AIG AGC TAC AAC TIG CTT GGA TTC
 1

110
 GAG GAG AGG GTG GGA GAA ACT CCC CTG ATG AAT GCG GAC TCC ATC TTG GCT GTG AAG AAA TAC TTG CGA AGA
 GLU GLU ARG VAL GLY GLU THR PRO LEU MET ASN ALA ASP SER ILE LEU ALA VAL LYS LYS TYR PHE ARG ARG
 LYS LEU GLU LYS GLU ASP PHE THR ARG GLY LYS LEU MET SER SER LEU HIS LEU LYS ARG TYR TYR GLY ARG
 AAA CTG GAG AAA GAA GAT TTC ACC AGG GGA AAA CTC ATG AGC AGT CTG CAC CTG AAA AGA TAT TAT GGG AGG
 110
 120

130
 ATC ACT CTC TAT TAT CTG ACA GAG AAG AAA TAC AGC CCT TGT GCC TGG GAG GAT GTT AGA GCA GAA ATC ATG AGA
 ILE THR LEU TYR LEU THR LEU THR GLU LYS LYS TYR SER PRO CYS ALA TRP GLU VAL VAL ARG ALA GLU ILE MET ARG
 ILE LEU HIS TYR LEU TYR LYS ALA LYS GLU TYR SER HIS CYS ALA TRP THR ILE VAL ARG VAL GLU ILE LEU ARG
 ATT CTG CAT TAC CTG AAG GCC AAG GAG TAC AGT CAC TGT GCC TGG ACC ATA GTC AGA GTG GAA ATC CTA AGG
 130
 140
 150

160
 TCC CTC TCT TTA TCA ACA AAC TTG CAA GAA AGA TTA AGG AGG AAG GAA TAA CAT CTG GTC CAA CAT GAA AAC
 SER LEU SER LEU SER THR ASN LEU GLN GLU ARG LEU ARG ARG LYS GLU
 166
 ASN PHE TYR PHE ILE ASN ARG LEU THR GLY TYR LEU ARG ASN
 AAC TTT TAC TTC ATT AAC AGA CTT ACA GGT TAC CTC CGA AAC TGA AGA TCT CCI AGC CTG TGC CTC TGG GAC
 160

AAT TCT TAT TGA CTC ATA CAC CAG GTC ACG CTT TCA TGA ATT CTG TCA TTT CAA AGA CTC TCA CCC CTG CTA

I GGA CAA TTG CTT CAA GCA TTC TTC AAC CAG CAG AIG CTG TTT AAG TGA TGA GCT AAT GTA

-TA ACT AIG ACC ATG CTG ATA AAC TGA TTT ATC TAT TTA AAT ATT TAT TTA ACT ATT CAT AAG ATT TAA ATT

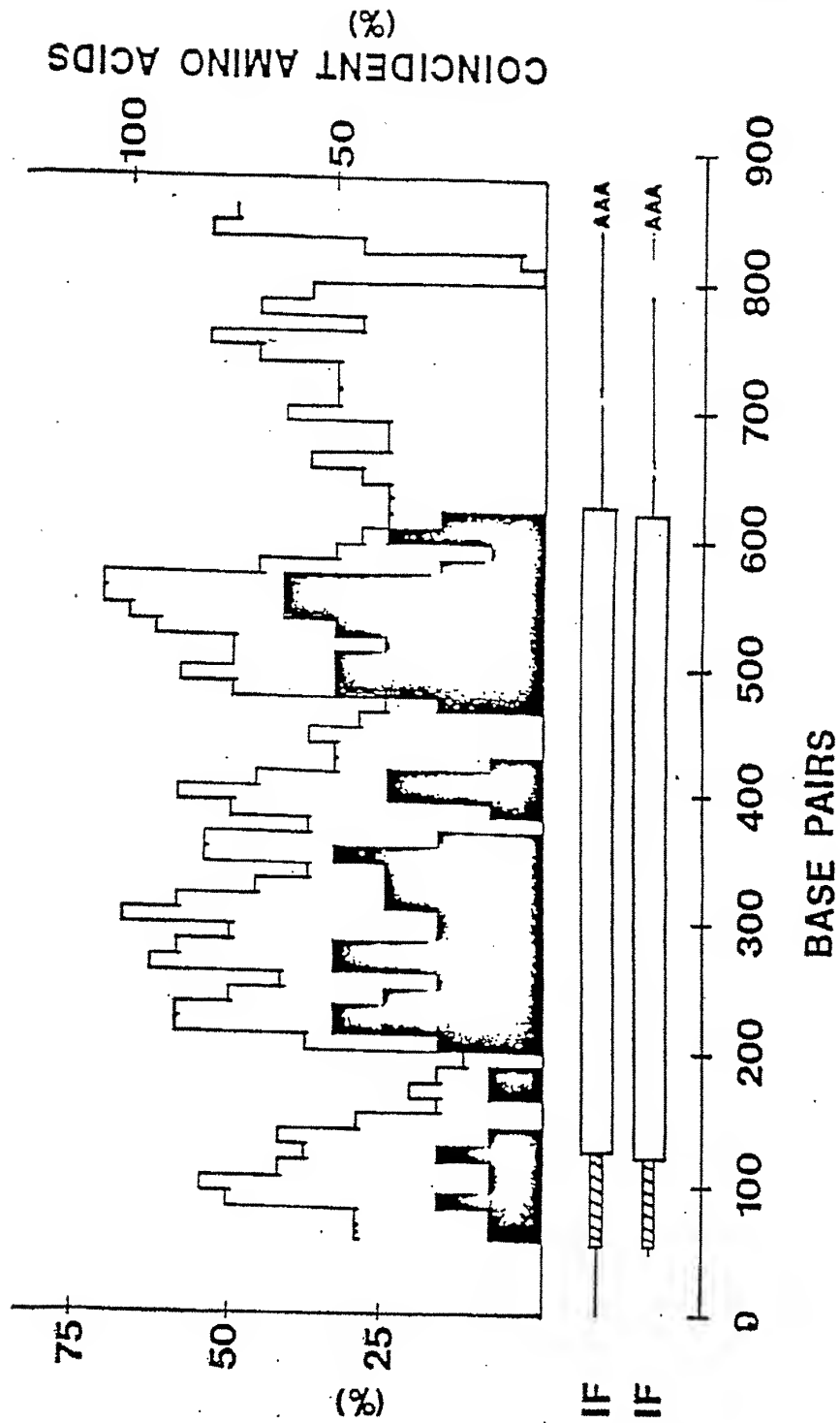
CIG CAT AIG AAA GGA CAC TAG AAG ATT IIG AAA TTT TTA TTA AAT TAT GAG TTA TTT TTA TTT AAT TAA ATT

ATT TTT GTT CAT ATA ACG TCA TGT GCA CCT TTA CAC TGT GGT TAG TGT AAT AAA ACA TGT TCC TTA TAT TTA

II AII IIG GAA AAI AAA TTA IIT IIG GIG CAA A-A

CTC AAA AAA A

CTC AAA A



TAB NN

PATENT OFFICE
JAPANESE GOVERNMENT

This is to certify that the annexed is a true
copy of the following application as filed with this Office.

DATE OF APPLICATION: October 30, 1979

APPLICATION NUMBER: 139289/79

APPLICANT(S): Juridical Foundation, Japanese
Foundation for Cancer Research

Dated this 31st day of October, 1980

Haruki Shimada
Director-General
Patent Office

Certified No. SHO 55-30579

Application for Patent

October 30, 1979

To: Mr. Yoshio Kawahara,
Director General of Patent Office

1. Title of the Invention
Novel recombinants having a gene which shows
complementarity to the human interferon messenger
RNA
2. Inventors
Name: Haruo Sugano (and other two)
Address: 4-8-13, Minami-Ogikubo, Suginami-ku, Tokyo
3. Applicant
Name: Juridical Foundation, Japanese Foundation
for Cancer Research
Hiroshi Anzai, Chairman Board of Directors
Address: 37-1, Kami-Ikebukuro 1-chome, Toshima-ku,
Tokyo
4. Attorney
Name: (6946) Junichi Sakata, Patent Lawyer
Phone: (984) 2023
Address: Hanabusa Bldg. 12-5, 2-chome, Minami-Ikebukuro,
Toshima-ku, Tokyo
Postal Code: 171
5. List of the Annexed Documents
 - 1) Specification one copy
 - 2) Power of Attorney one copy
 - 3) Duplicate of the Application one copy
6. Inventors other than the above
Name: Masami Muramatsu
Address: 4-21-6, Kotesashi-cho, Tokorozawa-shi,
Saitama-ken
Name: Tadatsugu Taniguchi
Address: 4-27-12, Tagara, Nerima-ku, Tokyo

- 1 -

Specification

1. Title of the Invention

Novel recombinants having a gene which shows complementarity to the human interferon messenger RNA.

2. Scope of Claim for Patent

(1) Novel recombinants that are made by inserting a DNA synthesized with human interferon messenger RNA as a template into a vector DNA, having a gene which shows complementarity to the human interferon messenger RNA.

(2) The novel recombinant in claim 1, wherein the recombinant is a recombinant plasmid DNA.

(3) The novel recombinant in claim 2, wherein the recombinant plasmid DNA is a recombinant Escherichia coli plasmid DNA.

3. Detailed Description of the Invention

This invention relates to novel recombinants having a gene which shows complementarity to the human interferon messenger RNA.

Interferon is a glycoprotein (molecular weight approx. 20,000) with antiviral activity, discovered by Isaacs and Lindenmann in 1957. Subsequent studies have revealed its antitumor activity besides its antiviral activity and hence a wide clinical application of this substance has gradually started. For instance, it is reported that this substance is effectively administered to various viral diseases, osteosarcoma and mammary carcinoma.

However, because of its high species-specificity, only the interferon derived from human cells can be used for human application. At present, the interferon which is being used for administration has a relative activity of about 10^6 (International units) per 1 mg, which corresponds to the purity of about 0.1 - 0.01%.

Further, the wide use of the interferon is impossible because of the difficulty of its mass-production. At present even for the interferon requirement for clinical tests (10^{13} units per year), the supply is only about 1%. For these reasons, development of technology to produce human interferon with a high purity easily and also in large quantities is desired.

The present inventors thought that it was the novel technique for producing interferon with ease and in a large quantity to insert a human interferon gene into a plasmid DNA (for instance plasmid DNA derived from Escherichia coli) with the recombinant DNA (deoxyribonucleic acid) technology. The inventors have completed this invention based on the thought.

That is, this invention relates to a novel recombinant, having a gene which shows complementarity to the human interferon messenger RNA, in which a DNA synthesized using the human interferon as a template is inserted into a vector DNA. The aim of this invention is to provide novel recombinants which grow and amplify in bacteria such as Escherichia coli and, as a result, can be used to produce human interferon in bacteria such as Escherichia coli.

The novel recombinant (or recombinant plasmid DNA) having a gene which shows complementarity to the human interferon messenger RNA has been obtained for the first time by the present inventors. The novel recombinant is a very useful substance which may be used for amplification in bacteria such as Escherichia coli and production of human interferon in large quantities and at low cost.

The novel recombinant of this invention can be obtained by the following procedure. First, cytoplasmic RNA is extracted from either human fibroblast, MG63 cells or others induced by poly(I)(C) which is a double-stranded RNA composed of polyinosinic acid and polycytidylic acid and is sold by the U.S. CALBIOCHEM Co. etc., or human leucocytes, lymphoblastoid cells, NAMALVA cells or others induced by Sendai virus, or lymphocytes induced by various

mitogens. From this RNA, the human interferon messenger RNA (hereinafter messenger RNA is referred to as mRNA) containing poly A (polyadenylic acid) is isolated. A double-stranded DNA is synthesized by reverse transcriptase, etc. with the mRNA portion having high interferon mRNA activity as the template. A recombinant is obtained by inserting the synthesized DNA into a vector DNA such as Escherichia coli plasmid DNA by the technique of recombinant DNA.

The process of producing the present recombinant is explained in detail below.

First, human fibroblasts may be obtained from fetus-derived foreskin, etc.

A small amount of interferon is added to the culture fluid of human fibroblasts to prime production of interferon by human fibroblasts, to which poly(I)(C) is added to induce the production of interferon. Cycloheximide is added simultaneously to increase the production of interferon (superinduction) and the level of interferon mRNA. At an appropriate time (e.g. 4 hours) after human fibroblasts are superinduced in this way, cells are collected and destroyed and nuclei are removed. Cytoplasmic total RNA is extracted with phenol, etc. RNA can also be extracted by destroying the whole cells, extracting both DNA and RNA with phenol, etc. and degrading and removing DNA with DNAase.

Further, RNA can also be extracted from MG63 cells which can be obtained from patients of osteosarcoma, etc. instead of human fibroblasts, by inducing as described above.

Furthermore, RNA can be induced in human leucocytes, lymphoblastoid cells which can be obtained from lymphoma patients or NAMALVA cells which can be obtained from lymphoma patients by incubation with Sendai virus (which can be obtained from medical schools), etc. and extracted therefrom as described above.

RNA can also be extracted from lymphocytes which are induced by various mitogens by the above described techniques.

The thus extracted RNA is dissolved in a salt solution with a high concentration (e.g. solutions of NaCl and KCl), and put on a column of oligo (dT) cellulose sold by P-L Biochemicals Co., USA to adsorb mRNA having poly(A) on the column. Elution is carried out with water, a salt solution with a low concentration such as 10 mM Tris-HCl buffer, or the like to isolate mRNA having poly(A).

The isolated mRNA is fractionated according to the difference mainly in molecular weight by sucrose density gradient centrifugation. Interferon mRNA activity in each fraction is checked by determining interferon activity (antiviral activity) of the protein which is synthesized in oocytes of African claw toad (Xenopus leavis) by micro-injecting a part of the fraction.

Then, a DNA complementary to the mRNA is synthesized in vitro by a reverse transcriptase which is obtained from avian myeloblastosis virus, using an mRNA having the highest specific activity fraction as the template.

The synthesis is carried out as follows.

An mRNA is reacted at an appropriate temperature (e.g. 37°C) for an appropriate period (e.g. 60 min.) with oligo(dT) sold by P-L Biochemicals Co., USA, $MgCl_2$ (e.g. 5 mM), NaCl (e.g. 30 mM), mercaptoethanol (e.g. 5 mM) and Tris-HCl buffer (e.g. pH 8.0, 40 mM) using a reverse transcriptase together with deoxyadenosine triphosphate (dATP), deoxythymidine triphosphate (dTTP), deoxyguanosine triphosphate (dGTP) and deoxycytidine triphosphate (dCTP) (e.g. 0.5 mM each) as substrates.

The thus obtained reaction product containing mainly DNA is subjected to deproteinization with phenol etc. and the template RNA is removed by alkali treatment. A double-stranded DNA is synthesized by a reverse transcriptase in a similar way as the synthesis of the DNA complementary to mRNA described above except that mRNA is replaced by DNA and oligo(dT) is excluded.

Incidentally, by using Escherichia coli DNA polymerase I which can be obtained from Escherichia coli MRE 600, etc. instead of reverse transcriptase, the same double-stranded DNA can be synthesized.

After the double-stranded DNA which is synthesized by the above described procedure is treated with Nuclease S₁ which can be obtained from Aspergillus oryzae in the presence of ZnCl₂ (e.g. 1 mM), etc., sodium acetate buffer (e.g. 0.1 M, pH 4.5), NaCl (e.g. 0.2 M), etc., deoxyadenine chains are formed at both 3' ends of the synthesized DNA by incubating with a terminal transferase (e.g. sold by Bethesda Research Laboratories, USA) purified from calf thymus in the presence of potassium cacodylate buffer (e.g. pH 7.6, 0.14 M), Tris (base) (e.g. 0.03 M), dithiothreitol (e.g. 0.1 mM), CoCl₂ (e.g. 1 mM) and dATP (e.g. 1 mM) at an appropriate temperature (e.g. 37°C) for an appropriate period (e.g. 20 min.).

On the other hand, a plasmid DNA which is used as a vector DNA, e.g. Escherichia coli plasmid pBR322 DNA [Gene vol. 2, p.95-113 (1977)], is cut at one site by treating with a restriction endonuclease EcoRI, which can be obtained from Escherichia coli RY13, etc. in the presence of Tris-HCl buffer (e.g. pH 7.5, 10 mM), MgCl₂ (e.g. 6 mM), NaCl (e.g. 0.1 M), mercaptoethanol (e.g. 6 mM), etc. and then treated with phage λ-derived exonuclease, which can be obtained from Escherichia coli W3102 (λ cI851 x 13), etc., in the presence of Na-glycine buffer (e.g. pH 9.5, 0.1 M), MgCl₂ (e.g. 5 mM), etc. Thereafter deoxythymidine chains are formed at both 3' ends in the same way as for the above-described synthesized double-stranded DNA by using dTTP instead of dATP.

Synthetic double-stranded DNA and plasmid DNA which are chain-elongated at both 3' ends as described above are incubated at an appropriate temperature for an appropriate period with Tris-HCl buffer (e.g. pH 7.5, 50 mM), NaCl (e.g. 0.1 M), EDTA (e.g. 5 mM), etc. and hybridized with hydrogen-bonds of adenine and thymine. Then, a transformable Escherichia coli strain, e.g. Escherichia coli X

1776 [Molecular Cloning of Recombinant DNA, Scott, W.A. & Werner, R. edited, Academic Press p.99-114 (1977)] is transformed with the hybridized DNA by the method of Enea et al. (J. Mol. Biol. vol.96, p.495-509, 1975) or the like.

In the novel recombinant plasmid DNA thus obtained, there exists a vector DNA gene, e.g. β -lactamase (enzyme that destroys ampicillin) gene, of Escherichia coli plasmid pBR322. Therefore, the transformed Escherichia coli shows resistance to ampicillin. The following technique is used to pick up a strain with a novel recombinant having a gene which shows complementarity to the human interferon messenger RNA among these ampicillin resistant strains.

First, [^{32}P] labelled DNA is synthesized with the RNA having interferon mRNA activity described above as a template and the DNA is hybridized with mRNA extracted, without induction by poly(I)(C) (therefore, this mRNA does not contain interferon mRNA), from the human fibroblast by incubating at a high temperature (e.g. 65°C) in a reaction mixture containing NaCl (e.g. 0.5 M), etc. Then, the hybridized DNA (Probe A) and non-hybridized DNA (Probe B) are separated by hydroxyapatite column chromatography. Next, filter-fixed DNAs of transformants are hybridized separately with Probe B or Probe A according to the technique of Grunstein-Hogness [Proc. Nat. Acad. Sci. USA, vol.72, p.3961-3965 (1975)] and strains having a DNA hybridizable with Probe B but not or hardly with Probe A are discerned by autoradiography.

Then, plasmid DNA is isolated from the thus discriminated strain and hybridized with mRNA having interferon mRNA activity by incubating at a high temperature (e.g. 53°C) in the presence of 80% (w/v) formamide, 0.4 M NaCl, etc. Since the mRNA hybridized with the plasmid DNA from the above-described strain does not pass through nitrocellulose filter and is trapped, this mRNA is eluted from the filter at a high temperature (e.g. 60°C) with a solution such as 90% (v/v) formamide and injected into oocytes of African claw toad and interferon activity is determined as described above.

When interferon activity is determined to be positive with this procedure, the DNA used for hybridization is concluded to be a DNA having a base sequence complementary to interferon mRNA and by this method, a recombinant plasmid DNA having a gene complementary to the human interferon mRNA can be discriminated.

Novel recombinants of this invention are very useful because they enable mass production of interferon in Escherichia coli or in eukaryotic cells which can be grown in a large scale.

The invention will be explained more in detail in the following example. However, this invention will not be restricted thereto.

Example:

After priming of human fibroblasts by overnight incubation with MEM culture medium containing interferon (25 U/ml), they were superinduced by adding 10 µg/ml of poly(I) (C) and 5 µg/ml of cycloheximide to the medium.

After 4 hours, 10⁹ superinduced human fibroblasts were destroyed by Teflon homogenizer in the presence of 0.3 % NP-40 and 50 µg/ml heparin in RSB buffer (10 mM Tris-HCl, pH 7.5; 10 mM NaCl; 1.5 mM MgCl₂). Nuclei were removed by centrifugation, and 9.6 mg of cytoplasmic RNA was obtained by extracting 3 times with phenol.

The cytoplasmic RNA was precipitated with ethanol, dissolved in 10 ml of 1 mM EDTA solution and incubated at 65°C for 2 minutes. 2.5 ml of a salt solution with a high concentration (0.5 M Tris-HCl, pH 7.5; 1 M NaCl; 50 mM EDTA) was added to the above solution and the mixture was put on an oligo(dT) cellulose column to adsorb mRNA containing poly(A). Elution was carried out with a salt solution with a low concentration (10 mM Tris-HCl, pH 7.5) and water to isolate 250 µg of mRNA containing poly(A).

The mRNA was precipitated with ethanol and dissolved in 0.5 ml of 1 mM EDTA solution. The solution was incubated at 65°C for 2 minutes, subjected to centrifugation through a 5-25% sucrose-density gradient (rotated

at 35,000 rpm using the SW40 rotor of Beckmann Co., U.S.A.) at 4°C for 16 hrs., and fractionated into 20 fractions.

The interferon mRNA activity of each of these fractions was determined as mentioned above. The result is shown in Table 1.

Table 1

| Fraction No. | Interferon Activity |
|--------------|---------------------|
| 9 | < 50 units/ml |
| 10 | 44 |
| 11 | 550 |
| 12 | 52 |

mRNA in Fraction No. 11 was approximately 5 µg. The mRNA and a reverse transcriptase were incubated at 37°C for an hour in 20 µl of a reaction mixture [5 µg mRNA; 1 µg oligo(dT); 8 units reverse transcriptase; 5 mM MgCl₂; 30 mM NaCl; 5 mM mercaptoethanol; 40 mM Tris-HCl, pH 8.0] and deproteinized with phenol. After RNA was removed by the treatment with 0.3 N NaOH at 37°C for 15 hours, the synthesized single-stranded DNA was incubated at 37°C in 20 µl of a reaction mixture [the same mixture as described above except that mRNA and oligo(dT) were excluded] for an hour to synthesize 1.5 µg of a double-stranded DNA.

The double-stranded DNA was treated with Nuclease S₁ in 50 µl of a reaction mixture (1.5 µg double-stranded DNA; 1 mM ZnCl₂; 0.1 M sodium acetate, pH 4.5; 0.2 M NaCl; 0.05 unit S₁) at 37°C for 30 minutes, and the enzyme was removed with phenol. The DNA was precipitated with ethanol and then treated with a terminal transferase in 20 µl of a reaction mixture [1.5 µg DNA; 0.14 M potassium cacodylate, pH 7.6; 0.03 M Tris (base); 0.1 mM dithiothreitol; 1 mM CoCl₂; 1 mM dATP; 1 unit terminal transferase] at 37°C for 20 minutes. Thus about 100 deoxyadenosine chains were formed at both 3' ends of the double-stranded DNA.

On the other hand, 10 µg of Escherichia coli plasmid pBR322 DNA was treated at 37°C for 2 hours with a

restriction endonuclease EcoRI in 100 μ l of a reaction mixture (10 mM Tris-HCl, pH 7.5; 6 mM $MgCl_2$; 0.1 M NaCl; 6 mM mercaptoethanol; 10 units EcoRI) leading to the cleavage at the only one cutting site in pBR322 DNA. The cut plasmid DNA was treated with an exonuclease derived from phage λ in 200 μ l of a reaction mixture (10 μ g DNA; 0.1 M Na-glycine, pH 9.5; 5 mM $MgCl_2$; 50 μ g/ml albumin; 17.5 units λ exonuclease) at 0°C for 90 minutes, and the enzyme was removed with phenol. The DNA was treated with a terminal transferase in 50 μ l of a reaction mixture (10 μ g DNA; 0.14 M potassium cacodylate, pH 7.6; 0.03 M Tris (base); 0.1 mM dithiothreitol; 1 mM $CoCl_2$; 1 mM dTTP; 2 units terminal transferase) at 37°C for 20 minutes. This led to the elongation of about 100 deoxythymidylate residues on both 3' ends of plasmid pBR322 DNA described above.

0.02 μ g of the synthesized double-stranded DNA thus obtained, and 0.1 μ g of the plasmid pBR322 DNA were incubated for hybridization in a solution containing 0.1 M NaCl, 50 mM Tris-HCl (pH 7.5) and 5 mM EDTA at 65°C for 2 minutes, at 45°C for one hour, at 37°C for one hour and at room temperature for one hour. Then, *Escherichia coli* x 1776 was transformed with the hybridized recombinant following the method of Enea et al.

About 4,000 ampicillin-resistant strains were isolated in this way. 3,600 resistant strains were chosen, and each DNA was fixed on two nitrocellulose filters (Grunstein-Hogness Method).

On the other hand, [^{32}P] labelled single-stranded DNA was synthesized (about 0.44 μ g, specific radioactivity approx. 6×10^8 c.p.m./ μ g) by a reverse transcriptase in the same way as that for single-stranded DNA mentioned above (dCTP was labelled with ^{32}P) using the interferon mRNA fraction (about 10 μ g) which had been extracted and partially purified as described above, as a template. The DNA was hybridized in 50 μ l of a reaction mixture (25 μ g mRNA; 0.45 μ g single-stranded DNA labelled with ^{32}P ; 0.5 M NaCl; 25 mM Pipes buffer, pH 6.5) at 65°C for 40 hours with

25 µg of mRNA extracted from human fibroblasts which had not been induced by poly(I)(C) by the same method as to extract poly(I)(C)-induced mRNA. The reaction mixture was put on a hydroxyapatite column, and elution is first carried out with 0.14 M phosphate buffer (pH 6.5) to eluate the single-stranded DNA, and then with 0.4 M phosphate buffer to eluate the DNA hybridized with RNA. As the result, the DNA (about 90 % of the whole) (Probe A) which hybridized with the mRNA mentioned above, and the DNA (about 10 % of the whole) (Probe B) which did not hybridize with it were isolated.

Each Probe was hybridized with the above DNA fixed on the nitrocellulose filter according to the Grunstein-Hogness method. By means of autoradiography four strains were isolated which react mainly to Probe B but little to Probe A.

Table 2 shows the extent of reaction of the DNAs from the four strains to each Probe as revealed by autoradiogram.

Table 2

| Ampicillin-resistant strains | Extent of Reaction of Probe with DNA in the strains | |
|------------------------------|---|---------|
| | Probe A | Probe B |
| # 319 | + | + |
| # 644 | + | + |
| # 746 | - | + |
| #3578 | + | + |

Plasmid DNA was isolated from cells of the four strains by the method of Currier and Nester [Analyt. Biochem. vol. 76, p.431-441 (1976)]. Then, these DNAs were hybridized with the interferon mRNA as follows;

First, 5 µg of plasmid DNA was incubated with restriction endonuclease Hind III which can be obtained from Haemophilus influenzae Rd in 50 µl of a reaction mixture (10 mM Tris-HCl, pH 7.5; 6 mM MgCl₂; 50 mM NaCl; 6 mM

mercaptoethanol; 5 units Hind III) at 37°C for 2 hours and thus cut with the restriction endonuclease. After deproteinization with phenol, the cut DNA was precipitated with ethanol and dissolved in 20 µl of 80 % (w/v) formamide. The solution was heat-denatured at 85°C for 10 minutes and was incubated in the solution consisting of 2.5 µg mRNA, 20 µl 80 % (w/v) formamide, 20 mM Pipes buffer (pH 6.5), 0.4 M NaCl and 5 mM EDTA, at 53°C. Four hours later the mixture was mixed with 0.4 ml of 3 x SSC (1 x SSC corresponds to 0.15 M NaCl, 0.015 M sodium citrate) at 0°C, and was filtered through a nitrocellulose filter (diameter : 1 cm, pore size : 0.45 µm) at a rate of about 0.5 ml per minute. After washing the filter with about 1.5 ml of 2 x SSC, the filter was immersed in a solution consisting of 0.6 ml of 90 % (v/v) formamide, 20 mM Pipes buffer, 0.1 % SDS (sodium dodecylsulfate) and 5 mM EDTA. The incubation of the filter at 60°C for 2 minutes and the removal of the solution were repeated 3 times and the RNA eluted from the nitrocellulose filter into the solution (1.8 ml) was precipitated with ethanol. mRNA containing poly(A) was isolated from the RNA by using oligo(dT) cellulose column chromatography, dissolved in a mixture of 3 µl of 10 mM Tris-HCl (pH 7.5) and 88 mM NaCl and injected into the oocytes of African claw toad. The interferon activity (antiviral activity) in the synthesized protein was determined.

Table 3 shows the interferon mRNA activity of the mRNA which has hybridized with the DNA derived from the four bacterial strains mentioned above.

Table 3

| Bacterial strain | Interferon mRNA activity (unit/ml) |
|------------------|------------------------------------|
| # 319 | 360 |
| # 644 | < 10 |
| # 746 | 15 |
| # 3578 | < 10 |
| pBR322DNA | < 10 |

Further experiment was carried out using strain #319 DNA.

5 µg of plasmid DNA obtained from strain #319 by the Currier and Nester method was cut with restriction endonuclease Hind III in the same way as mentioned above. The DNA and the recombinant plasmid βGpBR322 DNA (the vector was pBR322) (obtained from the Institute for Molecular Biology I of University of Zürich) containing rabbit β-globin gene, separately or as a mixture, were hybridized with the mixture of rabbit globin mRNA (obtained from rabbit red blood cells) (1 µg) and interferon mRNA (2.5 µg) obtained from human fibroblasts under the same conditions as mentioned above. The mRNA which formed hybrid was injected into the oocytes of an African claw toad. The oocytes were incubated for 15 hours in Barth's culture medium containing [³H] labelled histidine and [³H] labelled globin was isolated by acrylamide gel electrophoresis and determined quantitatively by fluorography. The interferon was determined by antiviral activity as described above. The synthesis of rabbit β-globin and the human interferon was determined in this way. The result is shown in Table 4.

Table 4

| D N A | Synthesized interferon activity | Amount of globin synthesized |
|-----------------------------|------------------------------------|---------------------------------|
| # 319 | 200 (units/ml) | - |
| βGpBR322 | 35 | + + + + |
| mixture of both plasmids | 160 | + + + |

From the result of this experiment it has been proved that DNA of #319 has DNA (the interferon gene) which forms a hybrid specifically with the interferon mRNA.

Applicant: Juridical Foundation, Japanese
Foundation for Cancer Research
Attorney: Junichi Sakata, Patent Lawyer

HARVARD UNIVERSITY
THE BIOLOGICAL LABORATORIES



February 1, 1980

16 DIVINITY AVENUE
CAMBRIDGE, MASSACHUSETTS 02138

Dr. Tadatsugu Taniguchi
Department of Biochemistry
Cancer Institute
Japanese Foundation for Cancer Research
Tokyo, Japan

Dear Dr. Taniguchi:

I enjoyed very much our phone conversation. My laboratory at Harvard University would be delighted to collaborate with you to achieve expression of your fibroblast interferon gene in E. coli. As you know, we have been working for some years developing methods that enable us to express efficiently eukaryotic (and prokaryotic) genes in E. coli. We have recently developed new procedures that are perfectly adapted to the problem of expressing the fibroblast interferon gene. I outline these procedures on enclosed separate pages. We believe that within a month or two we could construct plasmids that would direct synthesis of relatively high levels (approximately 10,000 molecules/cell) of two forms of fibroblast interferon. One of these would match exactly (in primary sequence) the molecule secreted by fibroblast cells, and the other would include the leader peptide. It would be interesting to compare the biological activities of these two molecules.

I imagine we might proceed in either of two ways. First, you might send us the clone. We would apply our methods to express it and immediately send to you those clones making the two forms of interferon mentioned above. We now have someone in the lab who can assay interferon activity and so we could see the matter through to this stage. Alternatively, you would be welcome to spend part or even all of the time required to do the work in my lab. (We are overcrowded, but we would manage somehow.) You might find it most interesting to come after we have done some of the necessary preliminary work with the clone. The exact procedure we use and hence the time involved, will depend upon the position of various restriction sites, etc. Still another possibility would be for you to come for the initial phase of the work.

It is important to note that I am writing to you solely as a member of Harvard University. I am not connected with any outside corporation (such as Biogen or Genentech) and we would work as scientific collaborators with no relation to members of those corporations. If it is necessary to enable you to visit us, I can pay for a part or all of your expenses.

Dr. Tadatsugu Taniguchi
February 1, 1980

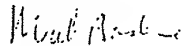
Page 2

I hope you share my feelings of enthusiastic urgency about this matter because, as you probably know, various groups in America (Genentech, Dupont, etc.) and elsewhere are working very hard to isolate and express the fibroblast interferon gene. Incidentally we have several people in the lab who are experts at DNA sequencing, and we could help you complete the sequence if that would expedite matters.

I hope to hear from you so that we can begin this most exciting collaboration very soon. Please feel free to telephone me collect. My numbers are:

617 495-2336
617 495-2329
617 868-8006

Yours truly,



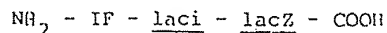
Mark Ptashne
Professor of Biochemistry and
Molecular Biology,
Harvard University

Expressing Eukaryotic Genes in E. Coli

The background to our current methods is found in the three enclosed papers (Backman and Ptashne, 1978 Cell 13,65; Roberts et al. '79a PNAS 76, 760; Roberts et al. '79b PNAS 76, 5596). The essential requirements are that the eukaryotic gene is transcribed efficiently and that translation of the mRNA begins at a specified AUG encoded by the eukaryotic gene. This results in production of a native (unfused) protein with methionine at its amino-terminus. In the case of fibroblast interferon, the first residue of the exported protein is methionine, and our method can be applied to produce either the entire protein including the leader peptide or only the smaller version that is secreted by the fibroblast cells.

The principle of our method is to create a DNA molecule encoding a "hybrid ribosome binding site", consisting of the ATG from the eukaryotic gene and the "Shine-Delgarno" sequence of the lacZ gene. When these two sequences are correctly joined the mRNA is translated efficiently beginning at the designated AUG. The first use of a hybrid ribosome binding site is described in Backman and Ptashne (1978). In Roberts et al. '79a we show that we can use a "portable promoter" to position a promoter (and a Shine-Delgarno sequence) at varying distances in front of a gene. In Roberts et al. '79b we show how this technique can be applied to express a eukaryotic gene (t). Our more recent methods greatly simplify and increase the efficiency of the procedures. These methods are unpublished and are outlined in the following paragraphs.

We have constructed three plasmids (series pLG) that bear the DNA encoding the amino-terminal portion of the laci gene fused to DNA encoding the carboxyl-terminal portion of lacZ. The carboxyl-terminal portion of β -galactosidase encoded by this DNA is enzymatically active. These plasmids have been modified by removing the 5' end of the laci gene and by inserting in the remaining portion of laci specific DNA linker fragments bearing restriction enzyme sites. When cut with the appropriate enzyme, each of the three plasmids exposes the i-Z fusion in a different translational reading frame. To express a eukaryotic gene (e.g., fibroblast IF), we proceed as follows. We first isolate a DNA fragment that encodes the amino-terminal portion of the IF gene. We then insert this DNA fragment into the appropriate pLG plasmid so that a fusion protein is encoded that would read:



Little or no β -gal would be synthesized, probably, because even if transcribed, the mRNA would not bear an efficient E. coli ribosome binding site. Next, we use the methods of Roberts et al. '79a and '79b to position the portable promoter, and its Shine-Delgarno sequence, at varying distances from the ATG of the IF gene. We transform cells and look for those colonies producing large amounts of β -galactosidase; these have the promoter (and Shine-Delgarno sequence) at optimal distances from the IF ATG, and are therefore producing relatively large amounts of the enzymatically-

Page 2

active fusion protein. It is now a simple matter, using recombination in vivo or in vitro, to replace the laci-2 portion of the fused gene with the 3' end of the IF gene, thereby producing a plasmid that directs synthesis of native IF. We can modify the procedure to use the ATG encoding either the methionine at the beginning of the leader or that at the beginning of the secreted protein. We have used this method to express β -globin and we are now writing a paper on this subject.

TAB OO

after SDS-polyacrylamide gel electrophoresis, the eluate should be centrifuged at 20,000 rev/min (Sorvall SS-34 rotor) for 20 minutes to remove particulate matter before dialysis. Coomassie blue staining of the gels to locate protein bands does not interfere with subsequent sequenator analysis.

New technologies such as the improved amino acid sequencing method described above lead to new research opportunities. With the greater sensitivity provided by this technique, we now can obtain amino acid sequence information on both proteins and peptides with submicrogram (picomole) quantities. This sensitivity should permit analysis of biomedically relevant molecules—such as the interferons—that can only be obtained in microgram quantities, and this ability opens possibilities for further study of these molecules. For example, knowledge of the amino acid sequence permits the synthesis of corresponding DNA probes and opens the possibility of

new strategies for isolating genes, such as those for interferons, that express low levels of messenger RNA's (8).

MICHAEL W. HUNKAPILLER

LEROY E. HOOD

Division of Biology, California Institute of Technology, Pasadena 91125

References and Notes

1. A. Isaacs and J. Lindenmann, *Proc. R. Soc. London Ser. B* **147**, 258 (1957).
2. T. Merigan, *N. Engl. J. Med.* **300**, 42 (1979).
3. E. Knight, Jr., M. W. Hunkapiller, B. D. Korant, R. W. F. Hardy, L. E. Hood, *Science* **207**, 525 (1980); K. C. Zoon, M. E. Smith, P. J. Bridgen, C. B. Anfinsen, M. W. Hunkapiller, L. E. Hood, *ibid.*, p. 527; H. Taira, R. J. Broeze, B. M. Jayaram, P. Lengyel, M. W. Hunkapiller, L. E. Hood, *ibid.*, p. 528.
4. P. Edman and G. Begg, *Eur. J. Biochem.* **1**, 80 (1967).
5. For a review of these methods, see M. Hunkapiller and L. Hood, *Biochemistry* **17**, 2124 (1978).
6. B. Wittmann-Liebold, *Hoppe-Seyler's Z. Physiol. Chem.* **354**, 1415 (1973).
7. N. Johnson, M. Hunkapiller, L. Hood, *Anal. Biochem.* **100**, 335 (1979).
8. B. Noyes, M. Mevarech, R. Stein, K. Agarwal, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1770 (1979).
9. Supported by a gift from the Ben Weingart Foundation.

29 November 1979

Human Fibroblast Interferon: Amino Acid Analysis and Amino Terminal Amino Acid Sequence

Abstract. The purification of human fibroblast interferon has been simplified to a two-step procedure consisting of affinity chromatography on Blue Sepharose and sodium dodecyl sulfate polyacrylamide gel electrophoresis. A preliminary amino acid composition and the sequence of the 13 amino-terminal residues of homogeneous interferon prepared by this method is reported.

Since the discovery of interferon, its purification and chemical characterization have been primary goals of interferon research. Although their attainment has been slow because of the small quantities of interferon proteins avail-

able, purification to homogeneity has now been achieved with some interferons. However, only microgram quantities have been available for characterization—human fibroblast interferon (1, 2), human lymphoblastoid interferon (3), human leukocyte interferon (4), mouse interferon (5)—and only limited structural information has been acquired (4, 6).

A thorough understanding at the molecular level of the numerous phenomena that are caused by interferon in cells in culture and in animals will not be possible until the elucidation of primary and secondary structures of the interferon proteins is achieved. This structural information will permit (i) comparison of amino acid sequences of interferons from various cell types and animal species, (ii) identification of the polypeptide segments involved in binding to interferon-specific cell-surface receptors, and (iii) chemical synthesis of interferons.

We now report an improved procedure for the purification of human fibroblast interferon that can be used to provide enough protein for structural studies.

Using the automated protein micro-sequencing technique described in (7), we have determined the sequence of the 13 amino acid residues at the amino terminus of the interferon prepared by this method. We also report a preliminary amino acid composition of the protein.

Human diploid fibroblast cells (FS-4) were cultured and interferon was produced (1). Interferon was assayed by a microtechnique (8) with vesicular stomatitis virus as the challenge virus. Interferon units are given in National Institutes of Health human fibroblast interferon units.

The crude interferon, 10 to 15 liters produced in the absence of serum, was made 1M in NaCl and passed at room temperature through a column (4 by 10 cm) of Blue Sepharose (Pharmacia, Inc.) equilibrated with 0.02M sodium phosphate buffer, pH 7.2, containing 1M NaCl. The interferon was retained whereas more than 95 percent of the total protein passed through the column. The interferon was eluted with a mixture of the column buffer and ethylene glycol (1:1), and each fraction was diluted immediately with 0.5 volume of the buffer (Fig. 1a). Fractions containing interferon activity were pooled, diluted with two volumes of the column buffer, and passed through a small (1 by 6 cm) Blue Sepharose column for concentration. The interferon was eluted as described above (Fig. 1b).

Fractions containing interferon were pooled, dialyzed against 1 mM tris-HCl,

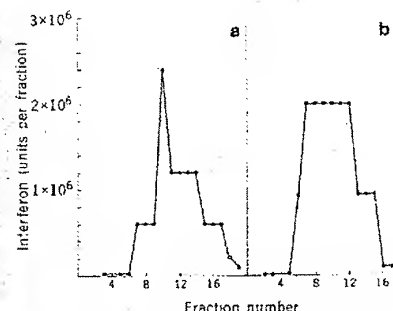


Fig. 1. (a) Fractionation of crude interferon on a large column of Blue Sepharose. Elution of interferon with 50 percent ethylene glycol in column buffer begins at fraction 1. (b) Small Blue Sepharose column. Fractions 7 to 17 in (a) were pooled, passed through the small column, and eluted with 50 percent ethylene glycol in column buffer (fractions 1 to 20).

Table 1. Amino acid composition of human fibroblast interferon.

| Amino acid | Composition | |
|------------|--------------|-----------------------------|
| | Mole percent | Residues per 20,000 daltons |
| Asp | 11.1 | 18.9 |
| Thr | 4.0 | 6.8 |
| Ser | 6.2 | 10.5 |
| Glu | 15.9 | 27.0 |
| Pro | 1.6 | 2.7 |
| Gly* | 4.6 | 7.8 |
| Ala | 5.9 | 10.0 |
| Cys† | 1.0 | 1.7 |
| Val | 3.5 | 6.0 |
| Met | 1.7 | 2.9 |
| Ile | 5.3 | 9.0 |
| Leu | 12.0 | 20.4 |
| Tyr | 4.4 | 7.5 |
| Phe | 5.5 | 9.4 |
| His | 2.9 | 4.9 |
| Lys | 6.8 | 11.6 |
| Arg | 6.4 | 10.9 |
| Trp‡ | 0.6 | 1.0 |

*Includes correction for free glycine present in unhydrolyzed protein. †Determined after performic acid oxidation. ‡Determined after hydrolysis with mercaptoethanesulfonic acid.

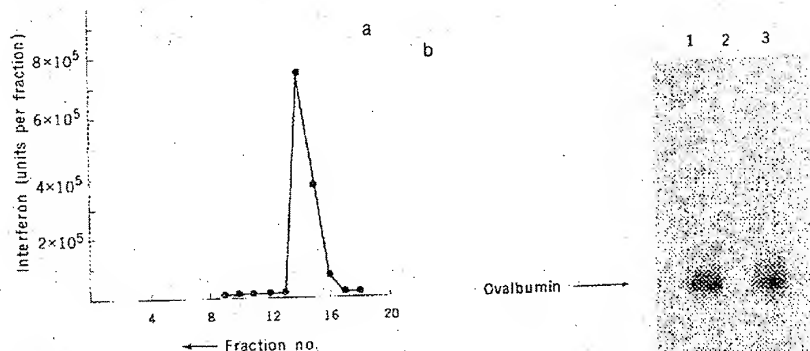


Fig. 2. (a) Preparative electrophoresis of interferon, activity profile. Fractions 6 to 15 in Fig. 1b were pooled, concentrated, and subjected to electrophoresis in a polyacrylamide slab gel, 0.75 mm thick. Fractions 14 and 15 were pooled and processed for amino acid sequencing. (b) Polyacrylamide slab gel, staining of proteins eluted from preparative gel in (a). Approximately 2 percent of the protein in fractions 14 and 15 (a) was subjected to electrophoresis and stained. Lanes 1 and 3, standard proteins; lane 2, interferon.

pH 6.8, containing 0.02 percent sodium dodecyl sulfate (SDS, Bio-Rad electrophoresis grade), and concentrated to dryness in a vacuum centrifuge. The interferon was then subjected to electrophoresis on a SDS-polyacrylamide slab gel and eluted (Fig. 2a). Fractions eluted from the gel were assayed for interferon activity (Fig. 2a). Approximately 0.2 μ g of interferon from the peak activity fraction was subjected to electrophoresis in this system again, and the gel was stained with Coomassie blue (Fig. 2b).

The preparative electrophoresis fractions containing interferon were pooled and centrifuged for 30 minutes at 30,000 rev/min at 4°C to remove polyacrylamide gel particles. The interferon solution was dialyzed first against 0.15M NaCl containing 0.1 percent SDS and then against 0.02 percent SDS. The dialyzed interferon was concentrated to dryness in a vacuum centrifuge.

This purification procedure is simpler and shorter than that described previously (1). Recoveries from the large Blue Sepharose column have ranged from 50 to 100 percent, and those from the small column approach 100 percent. The interferon (5×10^7 U/mg) eluted from these columns is stable for at least 4 weeks at 4°C in 1M NaCl, 35 percent ethylene glycol, pH 7.2. Recoveries of activity from the SDS gels have ranged from 5 to 20 percent, and specific activities of this protein have ranged from 2×10^8 to 8×10^8 U/mg. Accurate specific activities are difficult to determine, and

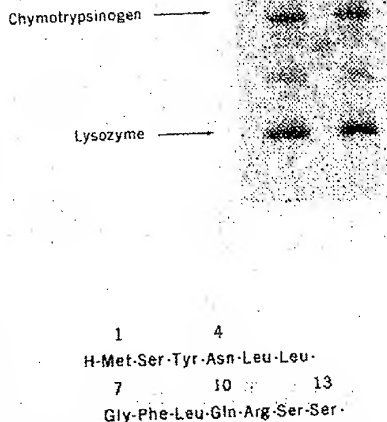


Fig. 3. The amino-terminal amino acid sequence of human fibroblast interferon.

two- to fourfold differences above 1×10^8 U/mg are probably not meaningful. Overall yields of purified interferon from 10- to 15-liter batches of crude material (5×10^7 to 7×10^7 total units, 8×10^4 U/mg) have averaged around 10 percent. This gives 5 to 10 μ g of homogeneous interferon.

Amino acid analysis on 1- to 2- μ g portions was performed on a Durrum D-500 amino acid analyzer (Table 1). Automated Edman degradation on 0.4- to 2- μ g portions of the purified interferon was performed on a spinning cup sequenator (7). Phenylthiohydantoin (Pth) amino acids were identified by high-performance liquid chromatography (HPLC) on a Du Pont Zorbax CN column (9).

The sequence of the 13 amino terminal amino acid residues of human fibroblast interferon was determined by this micro-sequencing technique (Fig. 3). Yields of Pth methionine at cycle 1 for three sequenator runs ranged from 60 to 100 percent (based on protein determination by amino acid analysis), and the sequenator repetitive cycle yields were 92 to 95 percent. Any unblocked minor peptide sequence present at > 5 percent of the reported sequence could have been detected by the methods used, but none has

been observed. This result coupled with the high initial Pth yields confirmed the homogeneity of the interferon polypeptide preparation.

Determining the amino acid sequence of a protein is essential in order to identify its active site and to understand the molecular mechanism of action. Comparison of structural features of interferons from different species and from different cell types within an animal will prove or disprove whether they are different proteins. If there is an active site common to all interferons, it should be identifiable by comparison of the amino acid sequences. Comparison of the amino terminal sequence reported here for human fibroblast interferon does not as yet reveal any apparent homology with the amino-terminal sequence reported for human lymphoblastoid interferon (10). However, there is limited homology (3 of 13 residues identical) with the 37,000-dalton mouse Ehrlich ascites cell interferon (11).

E. KNIGHT, JR.

Central Research & Development
Department, E. I. du Pont
de Nemours & Company,
Wilmington, Delaware 19898

M. W. HUNKAPILLER

Division of Biology, California Institute
of Technology, Pasadena 91109

B. D. KORANT, R. W. F. HARDY
Central Research & Development
Department, E. I. du Pont
de Nemours & Company

L. E. HOOD

Division of Biology,
California Institute of Technology

References and Notes

1. E. Knight, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **73**, 520 (1976).
2. W. Berthold, C. Tan, Y. H. Tan, *J. Biol. Chem.* **253**, 5206 (1978).
3. K. C. Zoon, M. E. Smith, P. J. Bridgen, D. zur Nedden, C. B. Anfinsen, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
4. M. Rubinstein, S. Rubinstein, P. C. Familletti, R. S. Miller, A. A. Walman, S. Pestka, *ibid.* **76**, 640 (1979).
5. E. Knight, Jr., *J. Biol. Chem.* **250**, 4139 (1975); Y. Yamamoto and Y. Kawade, *J. Gen. Virol.* **33**, 225 (1976); M. Kawakita, B. Cabrer, H. Taira, M. Rebello, E. Slattery, H. Weideli, P. Lengyel, *J. Biol. Chem.* **254**, 598 (1978); J. De Maeyer-Guignard, M. G. Tovey, I. Gresser, E. De Maeyer, *Nature (London)* **271**, 622 (1978).
6. B. Cabrer, H. Taira, R. J. Broeze, T. D. Kempe, K. Williams, E. Slattery, W. H. Konigsberg, P. Lengyel, *J. Biol. Chem.* **254**, 3681 (1979); Y. H. Tan, F. Barakat, W. Berthold, H. Smith-Johnson, C. Tan, *ibid.*, p. 8067.
7. M. W. Hunkapiller and L. E. Hood, *Science* **207**, 523 (1980).
8. J. A. Armstrong, *Appl. Microbiol.* **21**, 723 (1971).
9. N. J. Johnson, M. W. Hunkapiller, L. E. Hood, *Anal. Biochem.*, in press.
10. K. C. Zoon, M. E. Smith, P. J. Bridgen, C. B. Anfinsen, M. W. Hunkapiller, L. E. Hood, *Science* **207**, 527 (1980).
11. H. Taira, R. J. Broeze, B. M. Jayaram, P. Lengyel, M. W. Hunkapiller, L. E. Hood, *ibid.*, p. 528.

30 November 1979

SCIENCE, VOL. 207

TAB PP

Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 253, No. 24, Issue of December 25, pp. 8677-8679, 1978
Printed in U.S.A.

Nonglycosylated Mouse L Cell Interferon Produced by the Action of Tunicamycin*

(Received for publication, September 18, 1978)

Jun-ichi Fujisawa, Yoichiro Iwakura, and Yoshimi Kawade

From the Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606, Japan

SUMMARY

The effect of tunicamycin, a specific inhibitor of protein glycosylation, on interferon production by L cells induced by Newcastle disease virus was studied. The interferon yields in the presence of tunicamycin ranged from 20 to 100% of control in various experiments. The two molecular species of normal L cell interferon with 40,000 and 24,000 daltons were completely eliminated, and a new species of a smaller size (18,000 daltons) was produced. Radiolabeling of interferon with [³⁵S]methionine and [³H]glucosamine added to the induced cultures, as studied by immunoprecipitation and polyacrylamide gel electrophoresis, indicated that the two species of control interferon were both glycoproteins but tunicamycin interferon contained little or no sugar. In support of this was the loss of binding to lectin from *Wistaria floribunda*. A change in physical property of the molecule due to the absence of sugar was found in decreased heat stability in the presence of sodium dodecyl sulfate, although in its absence no significant difference from control interferon was observed. These results indicate that tunicamycin is an effective agent in producing nonglycosylated but fully active interferon molecules; the sugar moiety appears to be unnecessary for the antiviral activity and for secretion from cells.

Increasing evidence indicates that interferons are glycoproteins (1-5). As with many other glycoproteins, the functional role of the carbohydrate remains largely obscure. Dispensability of at least part of the carbohydrate moiety for the antiviral activity was indicated by glycosidase treatment of human interferons (3). Molecular heterogeneity of some interferons appears to be due to variations in their carbohydrate moieties. Removal of the carbohydrate by treatment of interferon with glycosidases or periodate, or by inhibition of glycosylation during its biosynthesis, resulted in reduced charge heterogeneity, as well as in decreased molecular weights (3-5). But these methods may not completely remove the carbohydrate, or they may entail inactivation and some side effects on the molecule.

To clarify the role of the carbohydrate moiety of interferon, we have examined the effect of tunicamycin on production of mouse L cell interferon induced by Newcastle disease virus. This antibiotic specifically inhibits formation of the *N*-acetyl-

glucosamine-lipid intermediate involved in *N*-glycosidic linking of core oligosaccharides to protein (6-8), and has been successfully used to inhibit glycosylation of various secretory and viral membrane proteins (9-14). A previous study in this laboratory indicated the glycoprotein nature of L cell interferon by its binding to some plant lectins (15). In this communication, we report that inhibition of glycosylation by tunicamycin eliminates interferon molecules of normal size and forces molecules of a smaller size containing little or no sugar but with full antiviral activity to appear.

MATERIALS AND METHODS

Interferon Induction—Monolayers of L cells were grown in Eagle's minimum essential medium supplemented with 5% calf serum on 30-mm plastic tissue culture dishes (Falcon), and infected with Newcastle disease virus at a multiplicity of 30 to 40 plaque-forming units per cell (16, 17). After an adsorption period of 1 h at room temperature, the inoculum was removed, and 0.6 ml of Earle's balanced salt solution containing L-glutamine (100 µg/ml) and kanamycin (200 µg/ml) was added to each dish. After 20 h at 37°C, the culture fluid was collected and clarified by centrifugation. To inactivate the virus, it was incubated with 0.5% Triton X-100 at 37°C for 10 min.

Tunicamycin, generously supplied by Professor G. Tamura of the University of Tokyo, was added after the virus adsorption process, and was present throughout the incubation period.

For double labeling of interferon, the culture fluid was replaced, at 7 h after infection (by which time little interferon was produced), by 0.6 ml of fresh medium with the glucose content reduced to 0.2 g/liter, containing 50 µCi of L-[³⁵S]methionine (1.98 Ci/mmol) and 120 µCi of [³H]glucosamine (28.8 µCi/mmol) (New England Nuclear, Boston). The fluid was harvested and processed as above.

Purification of Radioactive Interferon—Rabbit anti-interferon serum of high specificity, raised against partially purified L cell interferon (18), and absorbed by various conceivable impurities (19), was used to purify the radiolabeled interferon, as will be described in detail elsewhere.¹ To the culture fluid from each dish were added 25 µl of the antiserum, and the mixture was incubated at 37°C for 15 min and then at 4°C overnight. Goat anti-rabbit γ-globulin (Fujizoki Pharmaceutical Co., Tokyo), 17.5 µl, was added, and after incubation at 37°C for 30 min and at 4°C overnight, the precipitate formed was collected by centrifugation at 18,000 × *g* for 20 min. After washing twice with 0.2 ml of NaCl/P² containing 0.5% Triton X-100 and once with NaCl/P² alone, the precipitate was dispersed by sonication in 0.2 ml of 20 mM sodium phosphate buffer at pH 7.1 containing 0.15% SDS, 10 µM dithiothreitol, 2 mM EDTA, and 10% glycerol, and incubated at 70°C for 30 min. Then, 0.2 ml of NaCl/P² was added and 0.4 ml of 10% trichloroacetic acid was added at 4°C. The precipitate was collected by centrifugation and washed with 0.5 ml of 5% trichloroacetic acid at 4°C three times and with 0.5 ml of acetone at -10°C three times to remove trichloroacetic acid.

Polyacrylamide Gel Electrophoresis—SDS-polyacrylamide gel electrophoresis was performed as previously described (19), using 12.5% gel to minimize the influence of the sugar moieties of glycoproteins on their mobility. The precipitate of purified radioactive interferon was dissolved in 50 mM sodium phosphate buffer, pH 7.1, containing 5% SDS, 0.1 mM dithiothreitol, 2 mM EDTA, and 10% glycerol, heated at 70°C for 30 min, and electrophoresed at 1 mA/gel for 3 h and at 4 mA for 12 h. The distributions of interferon and radioactivity were determined on 2-mm slices of the gel, each of which was extracted at 4°C overnight with 0.2 ml of Eagle's minimal essential medium containing 5% calf serum.

Interferon Assay—The antiviral activity of interferon was deter-

* This work was aided in part by a grant from the Japanese Ministry of Education. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ S. Yonehara, Y. Iwakura, and Y. Kawade, manuscript in preparation.

² The abbreviations used are: NaCl/P², phosphate-buffered saline (0.02 M sodium phosphate, pH 7.4, 0.14 M NaCl); SDS, sodium dodecyl sulfate.

mined by the 50% inhibition point of [^3H]uridine incorporation into vesicular stomatitis virus RNA in L cells as previously described (18, 20). The titers in this report are expressed in international units.

Lectin-containing Column—*Wistaria floribunda* agglutinin, a D-galactose-specific lectin, covalently coupled to Sepharose, was supplied by Dr. Yoko Yamamoto of this laboratory, and used as described by her (15).

RESULTS AND DISCUSSION

When virus-induced L cells were incubated in the presence of tunicamycin (0.5 to 10 $\mu\text{g}/\text{ml}$), the interferon yields in the culture fluid varied considerably from one experiment to another, ranging from 20 to 100% of control without tunicamycin. The reason for this variation is not clear, but regardless of whether or not the yield was reduced, a distinct change in molecular size was observed for interferon produced in the presence of tunicamycin. Thus, L cell interferon normally consists of a major component of 40,000 daltons (S interferon) and a minor one of 24,000 daltons (F interferon), as analyzed by SDS-polyacrylamide gel electrophoresis (18, 19). Treatment with tunicamycin (2 $\mu\text{g}/\text{ml}$) resulted in almost complete disappearance of both the S and F species and in formation of a new peak of interferon activity with 18,000 daltons, as shown in Fig. 1. We shall designate this species as T interferon.

To make clear whether S, F, and T are glycosylated or not, we examined incorporation into them of [^{35}S]methionine and [^3H]glucosamine added to the induced cultures. The labeled interferons were purified from the culture fluid by immunoprecipitation and analyzed by SDS-gel electrophoresis. In Fig. 1a, the two peaks of the antiviral activity in the control preparation were accompanied by the radioactivity of both [^{35}S]methionine and [^3H]glucosamine, indicating that both S and F interferons are glycoproteins. (The radioactive peak at about 45,000 daltons is probably an impurity, which was not always observed, especially when smaller amounts of the anti-interferon serum were used.) In contrast, the incorporation of [^3H]glucosamine into the interferon proteins was strongly suppressed relative to that of [^{35}S]methionine in tunicamycin-treated cultures, as shown in Fig. 1b. It is concluded that the small interferon molecules (T) produced in the presence of tunicamycin contain little or no carbohydrate. The lower apparent molecular weight of T is consistent with this. Havell *et al.* (5), using 2-deoxyglucose or glucosamine as glycosylation inhibitors, also observed production of smaller interferon molecules from human fibroblasts, but in their experiments, normal size molecules were not completely eliminated even with an inhibitor dose that strongly suppressed the interferon yields.

The binding properties of the interferons to a lectin from *Wistaria floribunda* also indicated impairment of glycosylation due to tunicamycin. As shown in Fig. 2, control interferon was bound partly to the lectin and eluted with D-galactose followed by a weak acid, as previously reported (15), but interferon produced in the presence of tunicamycin was not bound at all.

These results are in good harmony with the known mode of action of the antibiotic, and indicate that the carbohydrate moiety of interferon molecules is dispensable for the antiviral activity, a conclusion also derived using different means of reducing the sugar content (3–5). The data in Fig. 1 indicate that the specific activity of T interferon, based on the [^{35}S]methionine radioactivity, was not lower than that of S and F molecules. Thus, tunicamycin appears to be a useful agent, in its specificity and completeness of action, to produce nonglycosylated and fully active interferon molecules.

The antiviral activity of normal L cell interferon was essentially stable against heating at 70°C for an hour in the presence of SDS and a reducing agent. But, interferon produced in the

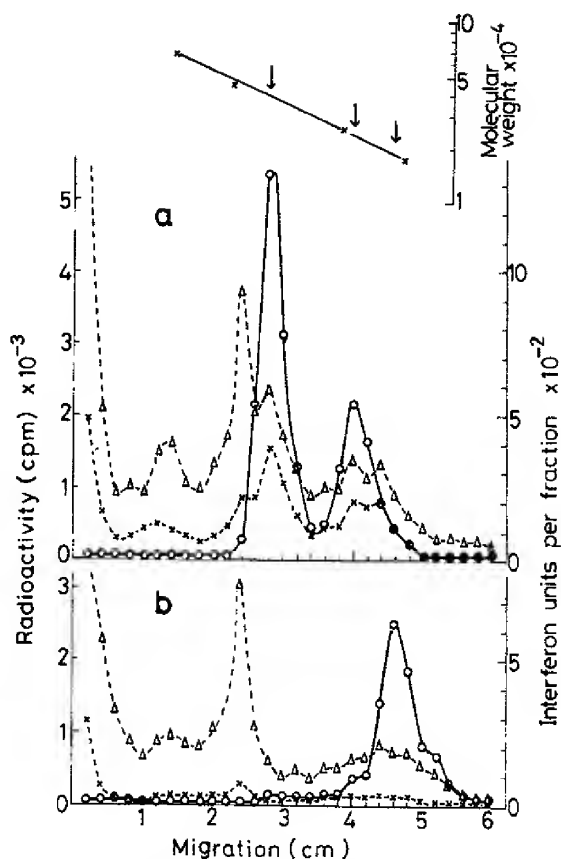


FIG. 1. SDS-polyacrylamide gel electrophoretic patterns of L cell interferon produced in the presence (b) and absence (a) of tunicamycin. Interferon samples labeled with [^3H]glucosamine and [^{35}S]methionine, purified from the culture fluids by immunoprecipitation (see "Materials and Methods"), were run for determination of the radioactivities. Portions of the same culture fluids, after incubation at 37°C for 20 min with added SDS (1%) and glycerol (20%), were electrophoresed for determination of interferon activity. a, control interferon. Recovery, 12% of interferon, 32% of ^3H , and 33% of ^{35}S radioactivities. b, interferon from cultures treated with tunicamycin (2 $\mu\text{g}/\text{ml}$). Recovery, 13% of interferon, 23% of ^3H , and 28% of ^{35}S radioactivities. O—O, interferon activity; x—x, ^3H radioactivity; and Δ—Δ, ^{35}S radioactivity. The apparent molecular weights were determined from the mobilities of the following reference proteins run in a parallel gel: bovine plasma albumin (68,000), ovalbumin (45,000), chymotrypsinogen A (25,000), and myoglobin (17,000).

presence of tunicamycin was inactivated by heating under the same conditions, as shown in Fig. 3a, indicating a substantial change in molecular characteristics due to the absence of sugar.

However, in the absence of SDS, the rates of inactivation at 45°C with or without 0.5% Triton X-100 were not significantly different between control and tunicamycin interferons, as shown in Fig. 3, b and c. Thus, it seems unlikely that thermal inactivation of nonglycosylated interferon in the culture fluid before being harvested was an important cause of reduced yields often observed in tunicamycin-treated cultures. As another possible cause of the yield reduction, we examined whether tunicamycin inhibited secretion of interferon to culture fluid without affecting its synthesis. In preliminary experiments, however, no appreciable accumulation of intracellular interferon was found, suggesting that the carbohydrate moiety of the molecules probably plays no essential role for secretion.

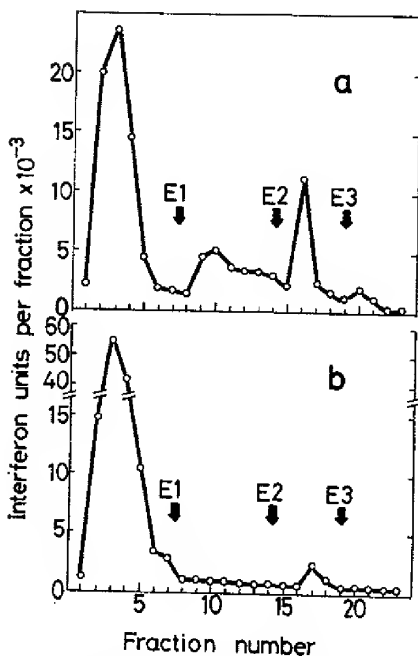


FIG. 2. Chromatography of interferons produced in the presence (b) and absence (a) of tunicamycin on *Wistaria floribunda* agglutinin coupled to Sepharose. Crude interferon samples, dialyzed against NaCl/P; containing 20% glycerol, were charged to a column (1 ml bed volume) containing 1 mg of the lectin, and the unadsorbed material was eluted with the same buffer. Further elution was done with: E1, 0.1 M D-galactose in NaCl/P; E2, 0.1 M acetic acid; and E3, 2 M NaCl, all containing 20% glycerol. Fraction volume, 1.5 ml. Recovery of interferon, 42% in a and 136% in b.

The structural relationship among S, F, and T is an intriguing question, and tunicamycin will be useful for clarifying it. We have found that purified S and F, upon tryptic digestion, produce a number of chromatographically identical peptides, suggesting the presence of common structure in their polypeptide moieties.³ The present results are consistent with the notion that S, F, and T differ from each other merely in the sugar moiety, and S and F are both formed by glycosylation of T. But other possibilities are by no means excluded. Further studies are in progress to elucidate the relationship among the different interferon species and the biological and structural roles of the sugar moiety.

Acknowledgments—We sincerely thank Professor Gakuzo Tamura of Department of Agricultural Chemistry, the University of Tokyo, for generously supplying tunicamycin and for his interest in this work, and Dr. Hiromi Mitsui, of Tokyo Metropolitan Institute of Medical Science for helpful discussions.

REFERENCES

- Schonne, E., Billiau, A., and DeSomer, P. (1970) in *International Symposium on Standardization of Interferon and Interferon Inducers* (Perkins, F. T., and Regamey, R. H., eds) pp. 61-68, Karger, Basel
- Dorner, F., Scriba, M., and Weil, R. (1973) *Proc. Natl. Acad. Sci. U. S. A.* **70**, 1981-1985

³ Y. Iwakura, S. Yonehara, and Y. Kawade, manuscript submitted for publication.

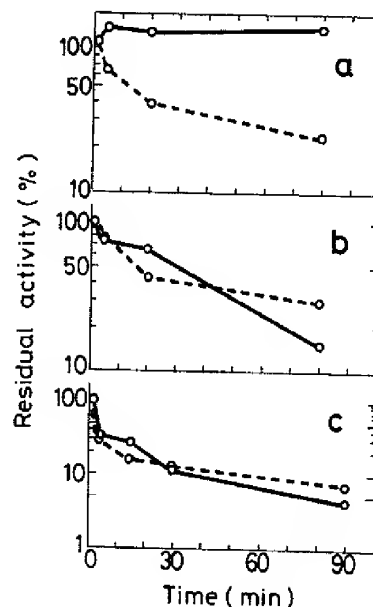


FIG. 3. Heat inactivation of interferon activity. ○—○, control interferon; ○—○, tunicamycin interferon. a, crude interferon samples (in Earle's balanced salt solution, containing 100 μg/ml of L-glutamine and 200 μg/ml of kanamycin) were made 0.1% in SDS and 10 μM in dithiothreitol, and heated at 70°C for the indicated times. The initial interferon titers were 5.0×10^5 and 1.7×10^5 units/ml, respectively, for control and tunicamycin interferons. b, the same crude interferon samples, made 0.5% in Triton X-100, were heated at 45°C. The initial titers were as above. c, crude interferon samples were heated at 45°C for the indicated times, and then Triton X-100 was added to 0.5% to inactivate the inducer virus. The initial titers were 1.7×10^5 and 0.7×10^5 units/ml, respectively, for control and tunicamycin interferons.

- Bose, S., Gurari-Rotman, D., Ruegg, U. T., Corley, L., and Anfinsen, C. B. (1976) *J. Biol. Chem.* **251**, 1659-1662
- Stewart, W. F., II, Lin, L. S., Wiranowska-Stewart, M., and Cantell, K. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 4200-4204
- Havell, E., Yamazaki, S., and Vilcek, J. (1977) *J. Biol. Chem.* **252**, 4425-4427
- Takatsuki, A., and Tamura, G. (1971) *J. Antibiot.* **24**, 232-238
- Takatsuki, A., Kohno, K., and Tamura, G. (1975) *Agric. Biol. Chem.* **39**, 2089-2091
- Tkacz, J. S., and Lampen, J. O. (1975) *Biochem. Biophys. Res. Commun.* **65**, 248-257
- Hickman, S., Kulczycki, A., Lynch, R. G., and Kornfeld, S. (1977) *J. Biol. Chem.* **252**, 4402-4408
- Duskin, D., and Bornstein, P. (1977) *J. Biol. Chem.* **252**, 955-962
- Olden, K., Pratt, R. M., and Yamada, K. M. (1978) *Cell* **13**, 461-473
- Schwarz, R. T., Rohrschneider, J. M., and Schmidt, M. F. G. (1976) *J. Virol.* **19**, 782-791
- Gibson, R., Leavitt, R., Kornfeld, S., and Schlesinger, S. (1978) *Cell* **13**, 671-679
- Nakamura, K., and Compans, R. W. (1978) *Virology* **84**, 303-319
- Yamamoto, Y. (1978) *J. Gen. Virol.*, in press
- Kawade, Y. (1973) *Jpn. J. Microbiol.* **17**, 129-140
- Yamamoto, Y., Tsukui, K., Ohwaki, M., and Kawade, Y. (1974) *J. Gen. Virol.* **23**, 23-32
- Yamamoto, Y., and Kawade, Y. (1976) *J. Gen. Virol.* **33**, 225-236
- Iwakura, Y., Yonehara, S., and Kawade, Y. (1978) *J. Biol. Chem.* **253**, 5074-5079
- Kawade, Y., Matsuzawa, T., Yamamoto, Y., Tsukui, K., and Iwakura, Y. (1976) *Annu. Rep. Inst. Virus. Res. Kyoto Univ.* **19**, 52-58

TAB QQ

Improved Methods for Maximizing Expression of a Cloned Gene: a Bacterium That Synthesizes Rabbit β -Globin

Leonard Guarente, Gail Lauer,
Thomas M. Roberts and Mark Ptashne
The Biological Laboratories
Harvard University
Cambridge, Massachusetts 02138

Summary

In this paper we describe a method for constructing *E. coli* plasmids that direct efficient expression of genes that encode eucaryotic or procaryotic proteins. No functional assays for the proteins are needed, and they are produced in their native, unfused state. The only requirement is that the genes be isolable without intervening sequences. We describe as an example the construction of a plasmid that directs the synthesis of about 10,000-15,000 monomers per cell of rabbit β -globin. The essential steps in a typical construction are as follows.

—A region of the gene encoding the amino-terminal portion of the protein is fused to DNA encoding an enzymatically active carboxy terminal fragment of β -galactosidase. The latter is carried on one of three plasmids designed to facilitate the fusion (the construction of these three plasmids is described in the Appendix).

—A "portable promoter" of the *lac* operon is placed at many positions in front of the fused gene using nucleases *in vitro*. Those promoter placements that elicit efficient expression of the fused gene are identified by the β -galactosidase activity that they express. (In the special case we describe, plasmids identified as directing efficient expression of β -globin were found to bear "hybrid" ribosome binding sites consisting of the Shine-Dalgarno sequence carried on the promoter fragment and the ATG of the β -globin gene.)

—The gene of interest is reconstituted intact, with the portable promoter in place, by recombination *in vitro* or *in vivo*.

Introduction

One current goal of recombinant DNA research is to maximize the expression of procaryotic and eucaryotic genes cloned on plasmids of the bacterium *E. coli*. Our methods, previously reported and extended in this paper, have been directed toward producing large amounts of the products in their native states, that is, unfused to other proteins. In this regard, our approach differs from that of some others who have described methods for producing proteins that are hybrids of procaryotic and eucaryotic amino acid sequences (for example see Fraser and Bruce, 1978; Mercereau-Pujalon et al., 1978; Seeberg et al., 1978;

Villa-Kamaroff et al., 1978). Our techniques were originally developed to produce large amounts of two regulatory proteins encoded by phage λ , the repressor and *cro* product (Backman and Ptashne, 1978; Roberts and Lauer, 1979; Roberts, Kacich and Ptashne, 1979a). More recently these methods have been extended to produce large amounts of other procaryotic regulatory proteins (G. Lauer, A. Poteete and R. Brent, unpublished data). We have also described the production of a native eucaryotic protein in *E. coli*, the SV40 t antigen (Roberts et al., 1979b).

The essential features of our previously described methods were based upon the following considerations. Expression of a cloned gene requires that it be efficiently transcribed and that the mRNA so produced be efficiently translated. Transcription depends upon a promoter (such as the *lac* promoter) positioned in front of the gene. Efficient translation of a message in bacteria depends at least in part upon the presence in the mRNA of a "ribosome binding site." This site is thought to consist of a sequence at least 3-9 bp long, called the Shine-Dalgarno (SD) sequence, followed 3-11 bp away by the AUG which encodes the amino terminal methionine of the protein (Shine and Dalgarno, 1975). The SD sequences are complementary to the 3' end of the 16S rRNA and probably promote binding by duplexing with that RNA (Steitz, 1979). The effects of sequences upstream from the SD sequence in the leader or downstream from the AUG in the gene are less clearly understood. It is known, for example, that the base following the AUG can influence binding to the ribosome (Taniguchi and Weissmann, 1978). We have shown that, in two cases, the sequence of the leader can have a dramatic effect on translation (Backman and Ptashne, 1978; Roberts et al., 1979a).

These considerations led to the following strategies. We used a DNA fragment (a "portable promoter") that encodes the promoter of the *lac* operon and 35 bp of *lac* mRNA. This mRNA includes the *lacZ* SD sequence and the next five nucleotides but does not include the AUG that encodes the amino terminal methionine of β -galactosidase. Using the appropriate nucleases, we positioned this portable promoter at many different places in front of the gene. For example, when positioned 9 or 11 bp in front of the small *t* gene of SV40 cloned on a plasmid, the plasmid directed the synthesis of relatively large amounts of small *t* protein. In each of these cases, we had formed a DNA fusion that encodes a hybrid ribosome binding site consisting of the SD sequence from the *lac* operon and the AUG encoded by the *t* gene. A similar hybrid ribosome binding site was used to express the λ repressor efficiently (Backman and Ptashne, 1978); other investigators used DNA synthetic techniques to express the human growth hormone (Goeddel et al., 1979). In

the case of the λ *cro* gene, the most efficient expression was achieved using the *cro* gene ribosome binding site. Even in the latter case, the distance of the promoter fragment from the gene profoundly influenced the level of protein produced (Roberts et al., 1979a).

An important limitation of our method described previously is the requirement for a functional or immunological assay to detect those clones that had the promoter positioned at the optimal distance from the initiator ATG of the gene to be expressed. In some cases, identifying such clones may be laborious or impossible. (For example, affinity-purified antiserum to rabbit β -globin cross-reacts in a solid phase radioimmune assay with at least one *E. coli* protein.) The chief advance we describe here enables us to express at high efficiency procaryotic or eucaryotic genes in the absence of assays for their gene products. The procedure requires knowledge of the approximate location of the ATG encoding the amino terminal methionine and the absence of intervening sequences in the gene to be expressed. In this paper, we describe these methods and show a particular application to the rabbit β -globin gene.

Results

Experimental Design

We wished to design a method whereby we could position our portable promoter fragment at an optimal distance from the 5' end of a gene in the absence of any assay for the gene product. We anticipate that in the case of eucaryotic genes, the optimal fusion would, in general, encode a hybrid ribosome binding site consisting of an SD sequence from the promoter fragment and the ATG from the gene. In the case of procaryotic genes, it may or may not be necessary to create such a hybrid ribosome binding site (compare Backman and Ptashne, 1978 and Roberts et al., 1979a). We have developed a method that uses *lacZ* gene fusions to enable us to identify readily clones bearing the promoter at the optimal distance from the gene. (See Bassford et al., 1978 and Ptashne et al., 1980 for examples of previous use of *lacZ* fusions.)

We have designed a series of plasmids (pLG200, pLG300 and pLG400; see Appendix) each of which contains a carboxy terminal region of the *lacZ* gene fused to a fragment of the *lacI* gene. The *lacZ* fragment, which comprises most of the gene, encodes a protein fragment that has β -galactosidase activity. This activity is not affected by additional protein sequences at its amino terminus (see below). The *I*-*Z* fused gene is not expressed in these plasmids because the *lacI* promoter and the end of the *I* gene have been deleted. Each of the plasmids bears a unique restriction site in the *lacI* portion of the *lacI*-*lacZ* fusion. When cut with the appropriate restriction enzyme, the *lacI*-*lacZ* hybrid gene is exposed in a

different translational reading frame in the three plasmids.

Figure 1 outlines the three steps required to express efficiently a gene (called *X*) using one of the plasmids described above and in the Appendix. First, we insert a fragment of DNA that encodes the amino terminal region of the *X* protein into one of the pLG plasmids. The particular pLG plasmid used is chosen so that the translational reading frames of *X* and *lacZ* are in register (Figure 1a). At this stage, the fused gene is, in general, neither transcribed nor translated. Second, using restriction digestion and appropriate nucleases, we position the portable promoter at varying distances from the ATG encoding the amino terminal methionine of the *X* protein (Figure 1b). This procedure is facilitated by the presence of a unique restriction site near the 5' end of the gene. If such a site is not present naturally, it can easily be introduced (Roberts et al., 1979a). Those fusions bearing the promoter at optimal distances from the amino terminal ATG are recognized as follows. *Lac*⁻ bacteria are transformed with plasmids bearing the portable promoter at varying distances from the gene. Bacterial colonies, now *Lac*⁺, bearing the plasmids that direct the synthesis of the enzymatically active hybrid protein, are recognized by their color on the appropriate indicator plates. The intensity of this color is a rough measure of the relative efficiency of production of the hybrid protein. The third step in the construction is to reconstitute intact *X* by substituting for the *lacI*-*lacZ* portion of the fusion the carboxy terminal region of *X* (Figure 1c). This can be done by using recombination in vitro or in vivo to generate the gene. The product of the reconstituted gene is readily visualized on gels using the maxi-cell technique (Sancar, Hack and Rupp, 1979).

A Bacterium Producing Rabbit β -Globin

Our source of the rabbit β -globin gene was the plasmid p β G1 (Maniatis et al., 1976), which encodes rabbit β -globin cloned as a cDNA copy (see Figure 2). We excised the gene using S1 nuclease and ligated Hind III linkers (Bahl et al., 1976; Sheller et al., 1977) onto the excised fragment. We then cut the fragment with Hind III and Bam HI (which recognizes a naturally occurring Bam HI site roughly two thirds of the way into the gene) and isolated a 330 bp amino terminal fragment containing 22 bp of the β -globin leader sequence and 301 bp of the β -globin coding sequence. This fragment was cloned into pBR322 (Figure 2a) to form pGL6 and re-excised on a larger fragment by cleavage with Pst I and Bam HI (Figure 2b). The latter fragment was then inserted into pLG300, producing a β -globin-*lacZ* fusion. This step of the construction joined two Pst I ends and two Bam HI ends that had been rendered flush (Backman, Ptashne and Gilbert, 1976). The preliminary cloning of the β -globin fragment into pBR322 facilitated the cloning into pLG300 because joining of the Pst I ends

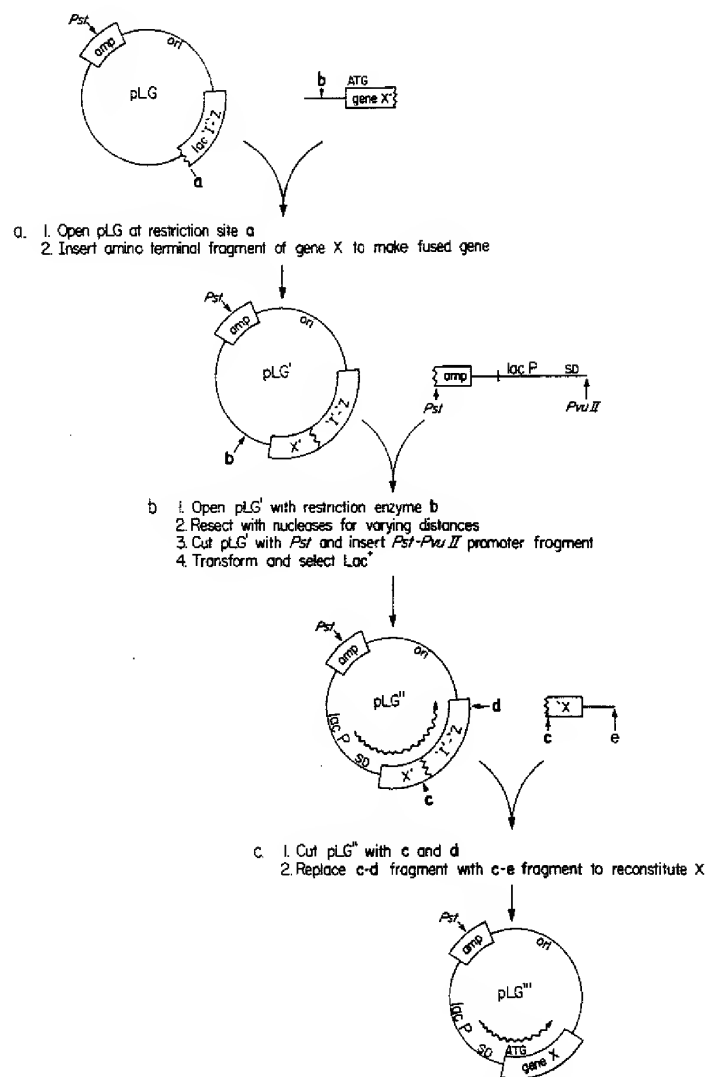


Figure 1. A General Method to Maximize Expression in *E. coli* of Any Prokaryotic or Eucaryotic Gene

(a) A fragment of DNA bearing the amino terminal region of gene X is inserted into restriction site a of a pLG plasmid, thereby fusing gene X to *lacI*-Z. In general, this fusion will occur by ligation of ends which have been rendered flush by DNA polymerase I (Backman et al., 1976).

(b) Plasmid pLG' bearing the fused gene is opened at a unique restriction site b which precedes the ATG of the fused gene. Resection and insertion of a portable promoter fragment that has a Shine-Dalgarno (SD) sequence is performed as described (Roberts et al., 1979a). Transformed clones that bear plasmids that direct the synthesis of high levels of β -galactosidase are identified as Lac⁺ colonies on the appropriate indicator plates.

(c) Gene X is reconstituted from plasmids that direct the synthesis of high levels of β -galactosidase as a fused product. This can be carried out, for example, by digesting the hybrid gene plasmid at any site (c) present in the gene X portion of the hybrid and at another site (d) in *lacZ*. Gene X is then reconstituted by the insertion of a DNA fragment that contains the carboxy terminal region of gene X extending from site c to a site (e) past the end of the gene. Gene reconstitution can also be achieved by in vivo genetic recombination.

Note: Plasmids are not drawn to scale in this or subsequent figures.

regenerates the *amp* gene. The structure of the resulting plasmid, pLG302, was confirmed by restriction enzyme analysis.

The portable promoter was positioned at varying distances from the β -globin ATG as follows (Figure 2c). pLG302 was cut with Hind III and resected with exonuclease III and S1 (Roberts et al., 1979a) in some cases and with Bal 31 (Gray et al., 1975) in others, and the reaction products were cleaved with *Pst*. An excess of the portable promoter fragment bounded on one end by a *Pst* site and on the other by a flush *PvuII* end was added, and the mixture was ligated. [This portable promoter is about 850 bp long. The *lac* SD sequence is 5 bp from the *PvuII* end and is adjacent to the *lac* promoter as previously described (Backman and Ptashne, 1978; Roberts et al., 1979a). Part of the *amp* gene is carried on the fragment, and joining of

the *Pst* ends of this fragment and pLG300 regenerates *amp* resistance. This promoter fragment is isolated from pGL101 (G. Lauer and M. Ptashne, unpublished results).] To identify those plasmids bearing the promoter at optimal distances from the β -globin ATG, a Lac⁻ strain of *E. coli* was transformed with the mixture and plated on lactose MacConkey indicator plates containing 15 μ g/ml ampicillin. Colonies producing β -galactosidase as indicated by their characteristic red color on the indicator plates were picked and assayed for the enzyme. Clones producing intermediate levels of β -galactosidase [about 300–700 units (Miller, 1972)] are pale red, and those producing high levels (~1000 units) are dark red under these conditions (see Figure 2 legend). (We describe below and in Figure 2d the reconstitution of the β -globin gene.) Clones that carried plasmids directing high levels of

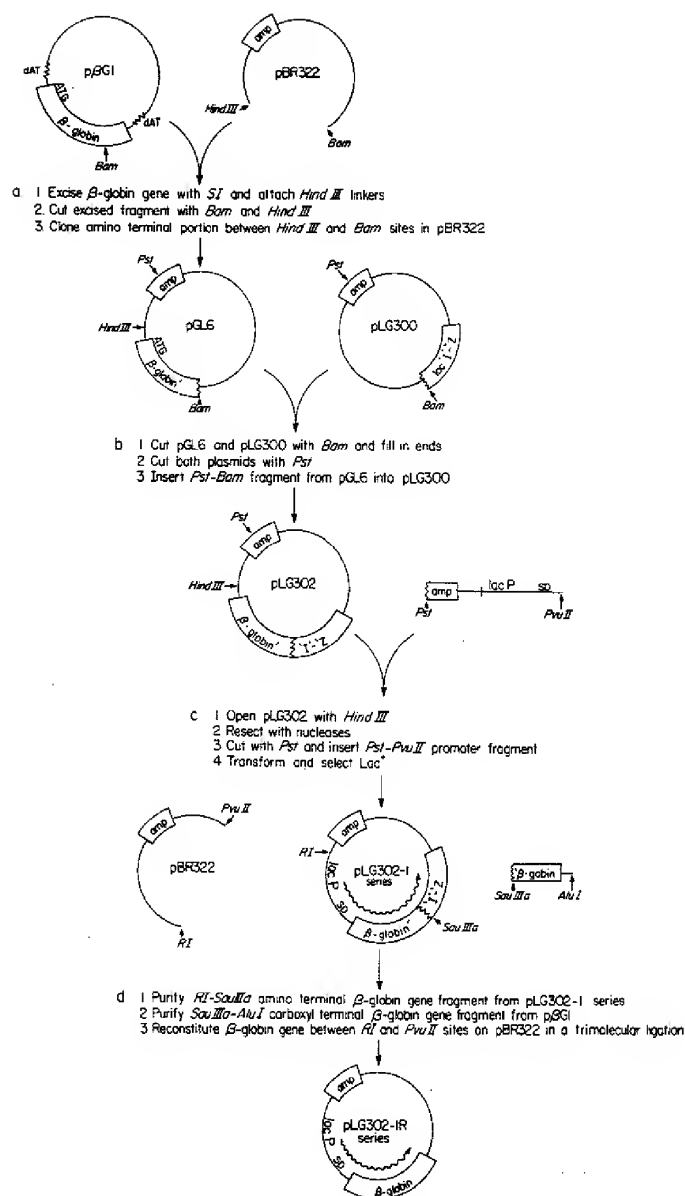


Figure 2. Construction of Plasmids That Direct the Synthesis of High Levels of Rabbit β -Globin in *E. coli*

(a) Plasmid p β G1 contains a cDNA fragment containing the rabbit β -globin gene with a 5' leader region of 56 nucleotides and a 95 nucleotide region to the 3' side of the gene (Efstratiadis, Kafatos and Maniatis, 1977), which was cloned by the dA.dT tailing method (Jackson, Symons and Berg, 1972; Maniatis et al., 1976). We excised this fragment by selective degradation with *S1* nuclease (Boeringer, 50,000 U/ml) as described (Hofstetler et al., 1976) and isolated a 525 bp fragment from a gel. *Hind* III linker oligonucleotides (Sheller et al., 1977) were ligated onto the ends of the fragment, and the molecule was cut with *Hind* III and *Bam* HI. *Bam* HI cuts 298 bp into the β -globin coding sequence. A 330 bp fragment containing a *Hind* III end, a 22 bp leader region (as determined by DNA sequencing) and 301 bp of the β -globin coding sequence were isolated from a gel and ligated with a purified pBR322 backbone fragment that had a *Bam* HI end and a *Hind* III end to yield pGL6.

(b) pGL6 and pLG300 (see Appendix) were separately digested with *Bam* HI and the ends were rendered flush (Backman et al., 1976), and then both were digested with *Pst* I. A purified fragment containing the amino terminal region of β -globin was ligated to a purified pLG300 fragment to construct the β -globin-*lacZ* fusion and regenerate the *amp* gene. The resulting plasmid is pLG302.

(c) pLG302 was opened by *Hind* III, resected with *Exo* III and *S1* (Roberts et al., 1979a) or *Bal* 31 (Gray et al., 1975) and then digested with *Pst* I. Resection was monitored by examining products at various times by electrophoresis on 3.5% acrylamide gels. A purified *lac* promoter fragment from pGL101 with a *Pst* I end and a *Pvu* II flush end was added in excess, and the mixture was ligated. An *E. coli* strain (NK5031) that bears a deletion of *lacZ* and synthesizes low levels of *lac* permease was transformed with the mixture and plated on lactose MacConkey indicator plates. Clones synthesizing ~300 units (Miller, 1972) of β -galactosidase are red on these plates. Plasmids from *Lac*⁺ clones were purified and the DNA sequence around the start of the gene was determined.

(d) A fragment with an *R1* and *Sau* 3A end containing the *lac* promoter and the amino terminal two thirds of the β -globin gene was purified from representative plasmids produced in the previous step. This fragment was ligated to a fragment with a *Sau* 3A end and an *Alu* I end containing the carboxy terminal third of β -globin (isolated from p β G1). Joining of the *Sau* 3A ends regenerates an intact β -globin gene. A pBR322 fragment with an *R1* end and a *Pvu* II end is also present in the ligation mix. A trimolecular ligation was performed which generated a pBR322 derivative containing an intact β -globin gene preceded by the *lac* promoter fragment.

β -galactosidase synthesis fell into two classes. One class was generated by *Bal* 31 or *Exo* III + *S1*, appeared at a frequency of about 3% among transformed clones and produced about 1200 units of β -galactosidase (for example, pLG302-2). The second class was generated only by *Exo* III + *S1*, appeared at a frequency of about 0.5% and produced about 2000 units of β -galactosidase (for example, pLG302-3). This corresponds to 10,000–20,000 monomers of

protein per cell (I. Zabin, personal communication). The DNA sequence was determined around the amino terminal coding region of six recovered plasmids. Three directed synthesis of high levels of β -galactosidase (two from the first and one from the second class), and three directed synthesis of lower levels of β -galactosidase. Figure 3 shows the relationship between the position of the portable promoter and the amount of β -galactosidase produced. The three hav-

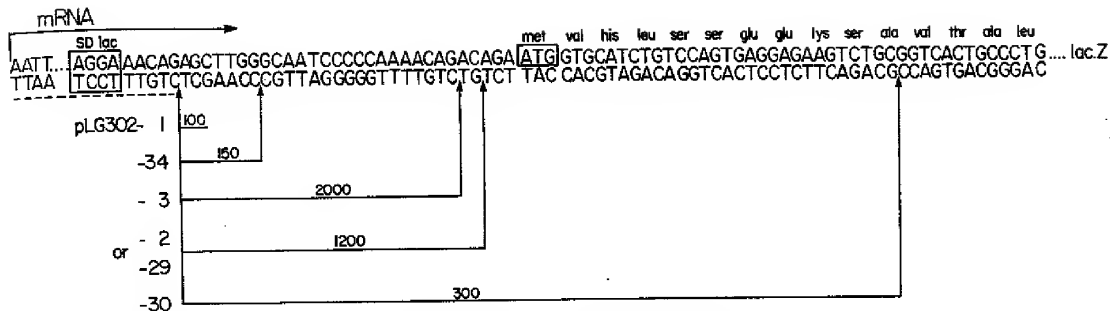


Figure 3. DNA Sequence around the Start of the Hybrid Gene after Various Positionings of the Portable Promoter Fragment

Depicted from left to right are the end of the *lac* promoter fragment (dashed line), the β -globin leader region, the first 15 codons of the β -globin gene and the distal *lacZ* fragment. In boxes are the Shine-Dalgarno (SD) sequence present on the promoter fragment and the ATG encoding the amino terminal methionine of β -globin. Arrows indicate the position of the *lac* promoter fragment in various plasmids, and the number of units of β -galactosidase directed by the plasmids is also shown. pLG302-1 was obtained by digesting pLG302 with Hind III, filling in the cohesive ends with DNA polymerase I and inserting the promoter fragment as described (Figure 2). The remaining plasmids were obtained by digesting pLG302 with Hind III, resecting with Exo III and S1 (pLG302-3) or Bal 31 (pLG302-2, -29, -30, -34) and inserting the promoter fragment. Each sequence was determined by first digesting the plasmid with RI and extending the exposed 3' ends with DNA polymerase and ^{32}P -labeled deoxyribonucleotide triphosphates. The DNAs were then cut with Pvu II, which cuts in the *lacI* interstitial fragment. The RI Pvu II fragment bearing the *lac* promoter and the β -globin gene fragment was isolated and sequenced.

ing the highest β -galactosidase levels (pLG302-2, -3 and -29) all had hybrid ribosome binding sites, and two of these had identical sequences. Of the other three, two had a large separation between the promoter fragment and the ATG, and one had been digested into the coding sequence. We have not determined where protein synthesis initiates in these latter three cases.

The hybrid proteins encoded by the derivatives of pLG302 identified as described above were visualized by polyacrylamide gel electrophoresis (not shown). For example, when compared with the parental strain, a strain harboring pLG302-2 contained an additional protein of approximate molecular weight 130,000. This is in good agreement with the predicted molecular weight of a β -globin- β -galactosidase hybrid protein. In addition, the level of hybrid protein seen on the gel was roughly that expected for a strain making 1200 units of β -galactosidase, assuming that the specific activity of the hybrid protein is roughly equivalent to that of wild-type β -galactosidase.

The intact globin gene carrying the *lac* promoter at different distances from the start of the gene was reconstructed from plasmids bearing hybrid genes as shown in Figure 2d. A fragment carrying the *lac* promoter and the 5' portion of β -globin was excised by RI and Sau 3A. (note that Sau 3A cuts at the same position at which β -globin was joined to pLG300.) This fragment was ligated to another containing the 3' end of the β -globin gene bounded by a Sau 3A site and an Alu site in the presence of a pBR322 fragment bounded by RI and Pvu II ends. This three-piece construction requires two sticky end (RI-RI and Sau 3A-Sau 3A) ligations and one flush end (Alu-Pvu II) ligation, and reconstitutes the β -globin gene. Following ligation, *amp*-resistant transformants were selected (Backman et al., 1976), and following plasmid

re isolation the construction was confirmed by restriction analysis. This β -globin reconstruction was performed with three different plasmids (pLG302-1, pLG302-2, pLG302-3; see Figure 3) which bore the promoter at different positions and which synthesized different amounts of the hybrid β -globin- β -galactosidase protein as determined by enzyme assay.

The production of native β -globin in cells bearing these three plasmids was demonstrated in two ways. First, the maxi-cell technique (Sancar et al., 1979) showed that all these plasmids directed synthesis of a protein the size of β -globin (Figure 4). The amount of the presumed β -globin made in each case was estimated by comparing, on polyacrylamide gels, the amount of radioactivity incorporated into the protein with that incorporated into a standard (see Figure 4 legend). The amount of β -globin made by the highest producer (pLG302-3) was estimated in this way to be about 10,000–15,000 monomers per cell, and that of the second highest producer (pLG302-2, -29) was estimated at about 5000–7500 monomers per cell. These values agree well with those predicted by the β -galactosidase activities directed by the parental fusion plasmids. The β -globin plasmid reconstructed from a low-level synthesizer of β -galactosidase (pLG302-1) produced no detectable β -globin. Second, a partial amino acid sequence of the protein encoded by one of the plasmids corresponded with that of authentic rabbit β -globin (Dayhoff, 1969). The protein was labeled with ^3H -leucine in one case and with ^{35}S -methionine in another, extracted from gels and subjected to 35 cycles of Edman degradation (Figure 5). In rabbit β -globin, there is no amino terminal methionine, and leucines are found at positions 3, 14, 28, 31, 32 In the labeled protein, leucines were found at positions 4, 15, 29, 32 and 33, and a methionine was found at position 1. This result shows

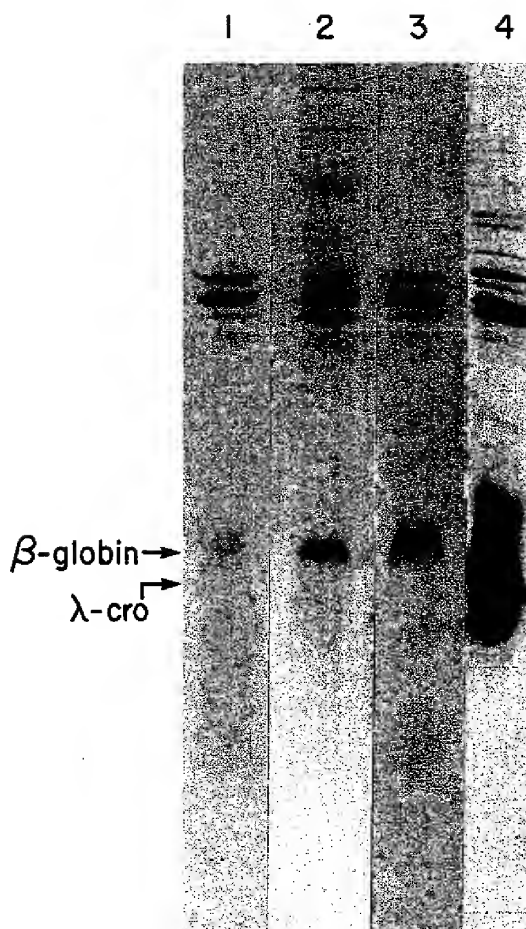
Cell
548

Figure 4. Polyacrylamide Gel Electrophoresis of Proteins Encoded by Plasmids Containing the β -Globin Gene and Various Placements of the Portable Promoter Fragment

Plasmid proteins were specifically labeled with ^3H -leucine by the maxi-cell technique (Sancar et al., 1979) in strain RB113 and analyzed by PAGE according to the method of Laemmli (1970) on 13.5% acrylamide gels. The position of a rabbit β -globin standard is shown. (1) pLG302-1R (reconstituted from pLG302-1); (2) pLG302-2R; (3) pLG302-3R; (4) pTR122. pTR122 makes about 40,000 monomers per cell of *cro* as determined by radioimmunoassay (Roberts et al., 1979a; T. M. Roberts, unpublished data). The number of monomers of β -globin made from various pLG derivatives was estimated by densitometry tracing of the autoradiogram and comparison with pTR122. The estimates are 5000–7500 monomers per cell for pLG302-2R and 10,000–15,000 monomers per cell for pLG302-3R.

that the protein is rabbit β -globin plus an amino terminal methionine which is not removed in *E. coli*.

Discussion

The methods described here and in a preceding paper (Roberts et al., 1979b) enable us to place a portable promoter fragment at many different distances in front of a gene and to recognize readily those placements

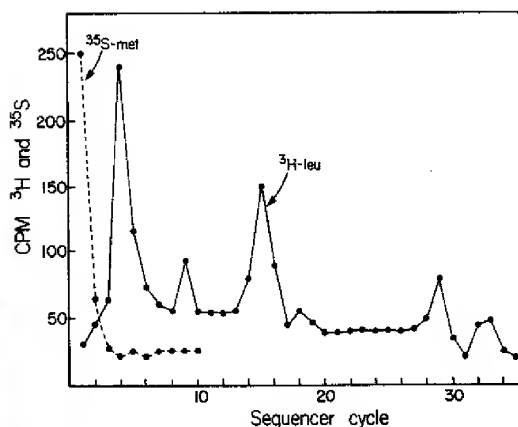


Figure 5. Position of Leucines and Methionines in the Amino Terminal Region of Bacterial β -Globin

After maxi-cell labeling with either ^3H -Leu or ^{35}S -Met, the 14 kilodalton protein encoded by pLG302-2R was extracted from the gel and analyzed by automated Edman degradation. The product of each cycle was counted in a liquid scintillation counter. The small peaks of ^3H -Leu at positions 9, 14 and 18 are presumably due to the presence of a contaminating protein in the gel eluent.

which elicit efficient transcription and translation of the gene. The method requires no assay for the gene product and can readily detect those plasmids bearing the promoter fragment at an optimal distance from the gene even if they occur rarely. For example, in the case we describe here, we positioned the portable promoter at varying distances in front of the ATG encoding the amino terminal methionine of rabbit β -globin. We immediately identified on the basis of β -galactosidase activity a small fraction (0.5%) that we suspected had an optimal promoter placement. In these cases, we confirmed that the plasmid directed the synthesis of high levels of rabbit β -globin when the gene was reconstituted. It is striking that our most efficient fusions selected only on the basis of β -galactosidase activity have positioned DNA encoding the Shine-Dalgarno (SD) sequence close to (8 and 10 bp) the ATG at the start of β -globin. These results may be taken as further support for the idea of a ribosome binding site as first proposed by Shine and Dalgarno (1975).

We do not yet have a systematic survey of the effect of varying the distance between the SD sequence and the ATG; nor do we know the effect of varying the sequence of the interstitial bases. Our strain producing the highest levels of β -globin makes about 10,000–15,000 molecules per cell. This is about an order of magnitude less than that expected on the basis of the known strength of the *lac* promoter and the number of copies of Col E1 plasmids in cells. (See Backman and Ptashne, 1978 for an argument that the *lac* promoter works as efficiently on a Col E1 plasmid as it does on the chromosome.) It remains to be seen whether sequences in the leader or within the coding

region of the gene are limiting expression. We have begun an analysis of mutations that increase synthesis of β -galactosidase from plasmids that encode hybrid genes (such as β -globin-*lacZ*). We do not yet know which portion of the plasmid sequence or protein synthetic machinery these mutations will change.

Our method will apparently work for expressing any procaryotic or eucaryotic protein beginning with its amino terminal methionine. In addition, our method may express shortened forms of the protein beginning at internal initiator codons. Finally, synthetic initiator codons can be attached within the gene to direct translation initiation, as has been reported in a previous case (Goeddel et al., 1979). In the two cases we have analyzed (β -globin and SV40 t antigen), the amino terminal methionine is not removed in *E. coli* (Roberts et al., 1979b).

Experimental Procedures

Bacterial Strains and Enzyme Assay

E. coli strains LG75 (F^- , *lacZ*_{YAS36}), NK5031 (F^- , *lacZ* Δ M5265, *suII*⁺, *nal*^R), LG90 (F^- , Δ *lacproXIII*), MM294 (*endo*⁺, *r_K*⁺, *m_K*⁺) and RB113 (F^- , *recA*13, *uvrA*6, *thi*1, *thr*1, *leuB*6, *proA*2, *argE*3, *his*4, *mtl*1, *xyI*5, *tsx*33, *stiA*31, *gal*2, *ara*14, *lacY*1) were used. Assay of β -galactosidase was performed as described (Miller, 1972), except that minimal media were supplemented with 0.3% casamino acids.

DNA Constructions

Restriction enzyme digestions and use of DNA polymerase I and T4 DNA ligase were as described (Backman et al., 1976), except that DNA polymerase I reactions were performed in the presence of 500 μ M dXTPs and ligations were performed in the presence of 600 μ M ATP. Use of exonuclease III and S1 was as described (Roberts and Lauer, 1979). To monitor digestion from the Hind III site of pLG302, an aliquot of DNA was subsequently cleaved with Hinf I, which cuts 130 bp into the β -globin coding sequence. Analysis of this digest on 3.5% acrylamide gels allowed the extent of exonuclease III digestion to be approximated to an accuracy of about 10 bp. Bal 31 exonuclease digestion was performed as described (Gray et al., 1975), except the reaction was carried out at 20°C and 0.2 M NaCl in a 60 μ l reaction volume containing 15 μ g of DNA. Under these conditions, the enzyme removed about 10 nucleotides per min. The reaction was stopped with phenol. DNA sequencing was performed by the method of Maxam and Gilbert (1977). Synthetic DNA linkers were from Collaborative Research; Hind III linkers used in the pGL6 construction were a gift from R. Sheller.

Radio-labeling of Proteins

Plasmid-encoded proteins were labeled in strain RB113 by the maxicell technique (Sancer et al., 1979). Use of ³⁵S-methionine was as described (Brent and Ptashne, 1980), while ³H-leucine labeling was performed as follows. A 10 ml culture of cells was grown, irradiated and shaken overnight as in the case of ³⁵S-methionine labeling. Cells were then spun and washed in M9 minimal medium containing all the amino acids except leucine. Cells were shaken in this medium for 1 hr at 37°C (typically in a 1 ml volume), and 100 μ Ci ³H-leucine (135 Ci/mmol) were added. Labeling with ³⁵S-methionine or ³H-leucine was performed for 1 hr, and cells were spun and resuspended in sample buffer (Laemmli, 1970). Labeled extracts were run on 13.5% acrylamide gels for analysis (Laemmli, 1970).

Automated Amino Acid Sequence Analysis

A 14 kilodalton protein was eluted from a 13.5% acrylamide gel (16,000 cpm of ³H-leucine-labeled material and, separately, 2000 cpm of ³⁵S-methionine-labeled material) and analyzed in a Beckman 890 B automated sequencer using a 0.1 M Quadrol program (Bruker

Margolies and Haber, 1975). The radioactivity released at each cycle was measured in a liquid scintillation counter. Unlabeled apomyoglobin (200 μ g) was added to the sample as an internal standard.

Appendix: Construction of Plasmids for Making *lacZ* Gene Fusions

Leonard Guarente and Mark Ptashne

Strategy

We wished to construct plasmids that would facilitate fusing a DNA fragment encoding the carboxy terminal portion of *lacZ* with any DNA fragment encoding an amino terminal portion of another gene. The hybrid protein was to express β -galactosidase activity. To do this, we exploited previous studies of *lacZ* gene fusion as follows. The first 41 amino acids of β -galactosidase may be removed without affecting enzymatic activity (Brickman et al., 1979). For example, one gene fusion was found to contain a large fragment of the *lacI* gene (355 of 360 codons) but was missing the first 23 codons of *lacZ* (Figure 6; Brake et al., 1978). The encoded hybrid protein possessed both *lac* repressor and β -galactosidase activities (Müller-Hill and Kania, 1974). We modified this *lacI*-*lacZ* hybrid gene so that DNA fragments encoding the amino terminal portions of other proteins could be inserted precisely into the *lacI* portion. If the translational reading frame of the inserted gene fragment is in register with the *lacI*-*lacZ* fragment, a tripartite gene is formed. If this gene is transcribed and if translation initiates at the amino terminal codon of the inserted gene fragment, a tripartite protein is made consisting of the product of the inserted gene fragment at the amino terminus, a sequence of amino acids encoded by *lacI* and an active β -galactosidase moiety at the carboxy terminus. Our plasmid constructions, described in detail below, consisted in outline of two steps. First, we constructed a plasmid containing the *lacI*-*lacZ* hybrid gene. Second, we replaced DNA encoding the promoter and amino terminus of the *lacI* gene with a variety of DNA linker oligonucleotides bearing sites sensitive to restriction enzymes. When cut with the appropriate enzymes, each of the plasmids exposes the *lacI*-*lacZ* gene in a different translational reading frame. This enables us to insert a fragment of DNA encoding the amino terminus of any protein to form a tripartite hybrid gene which encodes a protein with β -galactosidase activity.

Plasmid Construction

—(See Figure 6.) A DNA fragment containing an intact *lacI*-*lacZ* region was cloned into pBR322. As shown in the figure (a and b), this took place in two steps. β -galactosidase production from strains bearing the resulting plasmid, pLG2, was induced 15 fold by the addition of 10^{-3} M IPTG to the growth medium. IPTG (isopropylthiogalactoside) is a nonmetabolizable inducer of the *lac* operon. A DNA fragment including the end of *lacI* and the beginning of *lacZ* was deleted from pLG2, thereby forming a *lacI*-*lacZ* hybrid gene. The deletion was introduced by genetic recombination between a *lacZ*⁻ derivative of pLG2 and an F factor carrying the Δ 71-56-14 *lacI*-*lacZ* fusion (see Figure 6; Brake et al., 1978) (Figure 6c). Clones carrying pLG5 were constitutive for production of β -galactosidase, which is expressed by transcription initiating at the *lacI* promoter and translation initiating at the *lacI* amino terminal methionine codon.

—(See Figure 7.) Convenient restriction sites were introduced into the *lacI* portion of pLG5 as synthetic DNA linkers (Sheller et al., 1977). We introduced the 10 bp Hind III linker into a Pvu II site in *lacI* (Figure 7a). When the resulting plasmid, pLG200, is cut with Hind III and the cohesive ends are filled in with DNA polymerase, the DNA encoding a portion of *lacI* and the carboxy terminal region of *lacZ* are exposed in a particular translational reading frame. Insertion of an amino terminal gene fragment with a transcriptional initiation site and a translational initiation site in the appropriate reading frame with respect to the *lacZ* gene fragment will result in the synthesis of β -galactosidase as part of a tripartite polypeptide. To accommodate any of the three possible reading frames specified by a given amino terminal gene fragment, two additional derivatives of pLG200 were constructed. pLG200 was cut with Hind III, the cohesive ends were filled in with DNA polymerase (Backman et al., 1976) and the 10 bp

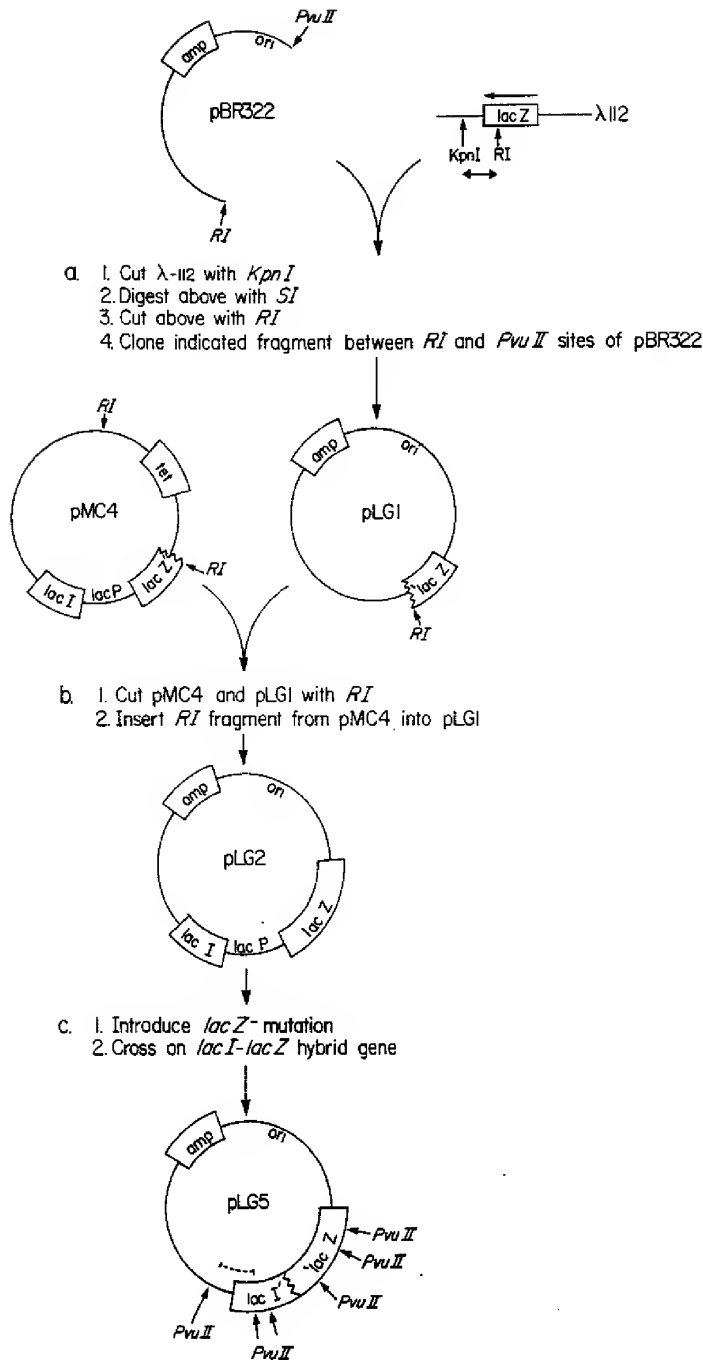


Figure 6. Construction of *lacZ* Fusion Plasmids

(a) Cloning a DNA fragment encoding the carboxy terminus of β -galactosidase. The *lacZ* transducing phage λ 112 (Maurer, Meyer and Ptashne, 1980) was digested with *Kpn*I, treated with *S*I nuclease to give blunt ends as described (Roberts and Lauer, 1979) and then treated with *Eco*RI. The fragment containing the carboxy terminus of *lacZ* was purified and ligated with a purified fragment of pBR322 DNA containing *R*I and *Pvu*II ends. The ligation mix was used to transform strain LG90 (*F*⁻, Δ *lac*) to ampicillin resistance (*Amp*^R). Plasmid DNA was isolated from several clones, and one containing the carboxy terminal fragment of *lacZ* (pLG1) was identified by analysis with restriction enzymes.

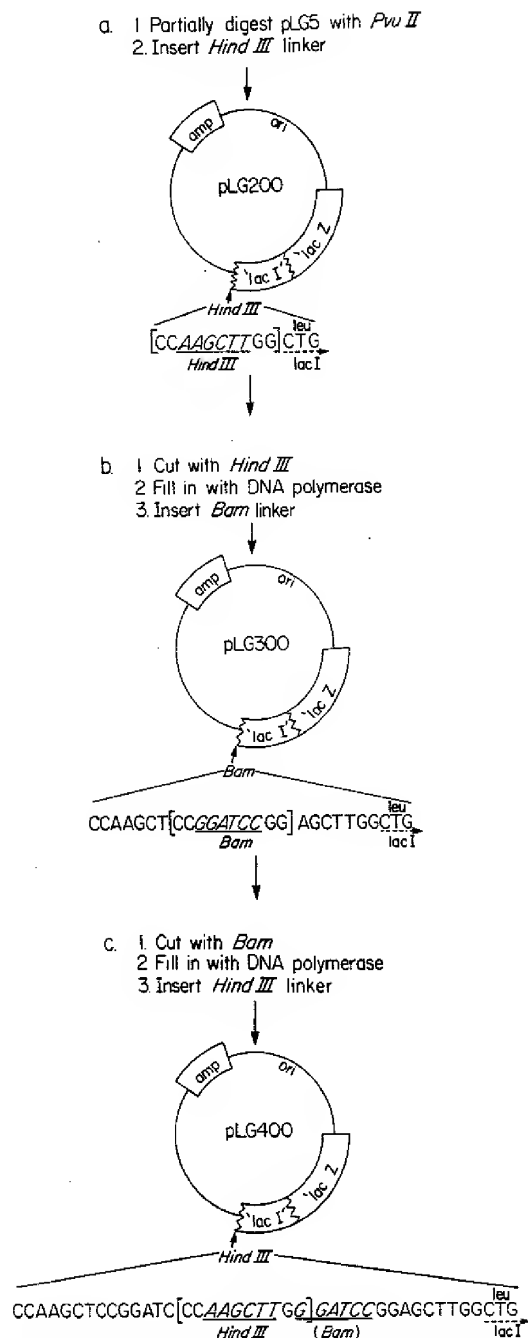
(b) Cloning complete *lacI* and *lacZ*. An *R*I fragment containing *lacI*, *lacP* and the amino terminal region of *lacZ* was purified from pMC4 (Calos, Johnsrud and Miller, 1978). This fragment was ligated with pLG1 DNA which had been cut with *R*I, and the ligation mix was used to transform LG90 to *Amp*^R on plates containing XG and 10^{-3} M IPTG. Blue colonies were purified, and plasmid DNA from one (pLG2) contained an intact *lacI*-*lacP*-*lacZ* region.

(c) Introduction of the *lacI*-*lacZ* fusion. A *lacZ*⁻ mutant of pLG2 was isolated by the phenotype of a colony bearing the plasmid on the indicator XG. (XG is a noninducing substrate for β -galactosidase that is hydrolyzed to form the blue dye indigo.) This mutant proved to have acquired an IS1 insertion in *lacZ* (data not shown). The *lacI*-*lacZ* hybrid gene, hybrid 71-56-14, was then crossed onto the *lacZ*⁻ plasmid by recombination in vivo. This was accomplished by introducing pLG2/*lacZ*⁻ into a strain bearing *F*'*lac* Δ 71-56-14. After many generations of growth, the plasmid was reisolated and introduced into a *LacZ*⁻ strain LG90. *Amp*^R/*lacZ*⁺ transformants were recognized on plates containing ampicillin and XG. The presence of the hybrid gene on a plasmid (pLG5) derived from one such clone was demonstrated by a 200 bp reduction in the size of a *Bcl*I fragment that spans the *lacI*-*lacZ* region. The hybrid gene has both *lac* repressor and β -galactosidase activities (Müller-Hill and Kania, 1974), and it is made in relatively large amounts because its promoter is the mutant form *I*⁰ (Müller-Hill, Crapo and Gilbert, 1968). The dashed *Pvu*II fragment is described in the legend to Figure 7a.

*Bam*HI linker was inserted to yield pLG300 (Figure 7b). This process destroys the *Hind*III site. Similarly, pLG300 was cut with *Bam*HI, the ends were filled in and a *Hind*III linker was inserted to give pLG400 (Figure 7c). By cutting pLG200, pLG300 and pLG400 with the appropriate restriction enzymes and filling in the cohesive ends, *lacZ*

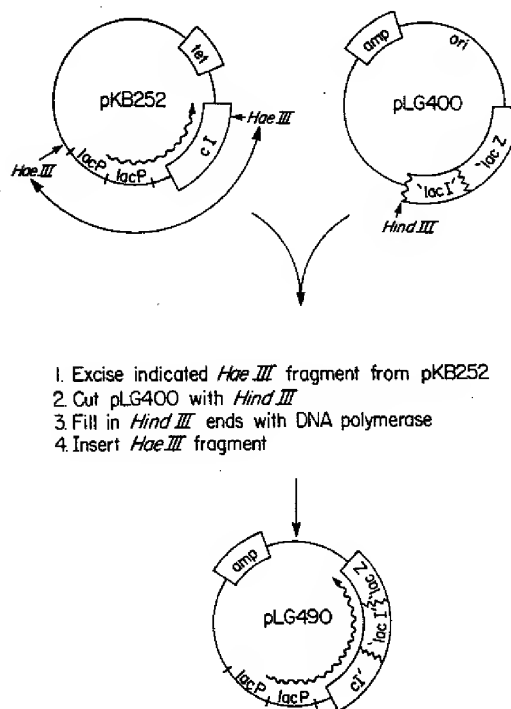
can be fused in frame to any given amino terminal gene fragment having flush ends.

To verify that the above manipulations created the predicted DNA sequence, we constructed a fusion of the *lacZ* gene fragment of pLG400 to a *Hae*III fragment of the *cl* gene of phage λ that encodes

Figure 7. Construction of *lacZ* Fusion Plasmids (Continued)

(a) Introduction of linkers. pLG5 was partially digested with *Pvu* II and religated in the presence of *Hind* III linkers. Following religation and transformation, a plasmid bearing a *Hind* III linker in place of the *Pvu* II fragment indicated in Figure 6 by a dashed line was recognized as follows. The ligated DNA was introduced into LG75 (F^- , *lacZam*_{YAC206}). In this strain, the presence of the *lac* repressor supplied by the plasmid results in a Melibiose⁻ phenotype because it represses the

all but the carboxy terminal six amino acids of λ repressor (Sauer, 1978) (Figure 8). The resulting plasmid, pLG490, conferred upon host clones immunity to phage λ and also directed the synthesis of β -galactosidase. These activities apparently reside on one hybrid protein with the λ repressor fragment at the amino terminus and β -galactosidase at the carboxy terminus. Clones bearing pLG490 produce a large protein of approximately the expected size (see Figure 9).

Figure 8. A *ci-lacZ* Fused Gene

A *Hae* III fragment containing the first 231 (of 237) codons of wild-type *cI* (Sauer, 1978) was isolated from pKB252 (Backman et al., 1976). This fragment also bears two *lac* promoters that direct transcription of the *cI* gene. The fragment was ligated with pLG400 DNA that had been cut with *Bam* HI and its ends rendered flush with DNA polymerase, and the ligation mix was used to transform LG100 (F^- , Δ *lac*) to Amp^r on plates containing XG. Blue colonies appeared at a frequency of about 10%, and these colonies also acquired immunity to phage λ . Plasmid DNA from one such clone had the anticipated structure.

chromosomal *lacY* gene. Removal of the *Pvu* II fragment was expected to render the plasmid incapable of directing the synthesis of repressor or β -galactosidase activities because the *lacI*⁺ promoter and 5' end of *lacI* would be missing. Still present, however, would be a region homologous to the portion of *lacZ* bearing the amber mutation on the chromosome. Amp^r clones were therefore screened for the following three phenotypes: LacZ⁻ (on XG); LacI⁻ (on Melibiose MacConkey indicator agar at 42°C); homology with the *lacZ* amber mutation (by recombination to LacZ⁺ on lactose MacConkey indicator agar). DNA was prepared from candidate clones and one, pLG200, had the correct structure. pLG300 and pLG400 were constructed as described in the text.

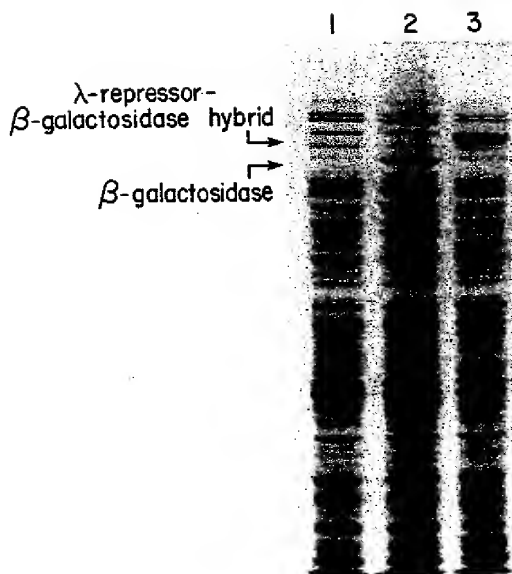
Cell
552

Figure 9. A λ Repressor- β -Galactosidase Hybrid Protein

Cells from saturated cultures were heated to 90°C for 3 min in sample buffer (Laemmli, 1970), and the extracts were electrophoresed in 7.5% acrylamide gels (Laemmli, 1970). The positions of β -galactosidase and the putative hybrid protein are indicated. Note that strain LG101 is not isogenic to LG100. (1) LG100 (F^- , Δlac); (2) LG101 (F^- , $\Delta lac/F' lac^+$, $lacZ^+$); (3) LG100/pLG490.

Acknowledgments

We thank R. Rogers Yocum for providing the protein sequence reported here.

L. G. is a postdoctoral fellow of the Jane Coffin Childs Memorial Fund for Medical Research. G. L. is a postdoctoral fellow of the NIH. This work was supported by grants from the NIH and NSF.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 10, 1980

References

- Backman, K. and Ptashne, M. (1978). Maximizing gene expression on a plasmid using recombination in vitro. *Cell* 13, 65-71.
- Backman, K., Ptashne, M. and Gilbert, W. (1976). Construction of plasmids carrying the *cl* gene of bacteriophage λ . *Proc. Nat. Acad. Sci. USA* 73, 4714-4718.
- Bahl, C. P., Mariani, K. J., Wu, R., Stawinski, J. and Narang, S. A. (1976). A general method for inserting specific DNA sequences into cloning vehicles. *Gene* 1, 81-92.
- Bassford, P., Beckwith, J., Berman, M., Brickman, E., Casadaban, M., Guarente, L., Saint-Girons, I., Sarthy, A., Schwartz, M., Shuman, H. and Silhavy, T. (1978). Genetic fusions of the *lac* operon: a new approach to the study of biological processes. In *The Operon*, J. Miller and W. Reznikoff, eds. (New York: Cold Spring Harbor Laboratory), pp. 245-262.
- Brake, A., Fowler, A., Zabin, I., Kania, J. and Müller-Hill, B. (1978).

β -Galactosidase chimeras: primary structure of a *lac* repressor- β -galactosidase protein. *Proc. Nat. Acad. Sci. USA* 75, 4824-4827.

Brauer, A., Margolles, M. and Haber, E. (1975). The application of 0.1 M Quadrol to the microsequence of proteins and the sequence of tryptic peptides. *Biochemistry* 14, 3029-3035.

Brent, R. and Ptashne, M. (1980). The wild-type *lexA* gene product represses the *lexA* promoter. *Proc. Nat. Acad. Sci. USA*, 77, 1932-1936.

Brickman, E., Silhavy, T., Bassford, P., Shuman, H. and Beckwith, J. (1979). Sites within gene *lacZ* of *E. coli* for formation of active hybrid β -galactosidase molecules. *J. Bacteriol.* 139, 13-18.

Calos, M. P., Johnsrud, L. and Miller, J. H. (1978). DNA sequence at the integration sites of the insertion element IS1. *Cell* 13, 411-418.

Dayhoff, M. (1969). *Atlas of Protein Sequence and Structure*, 4 (Silver Spring, Maryland: National Biomedical Research Foundation).

Efstratiadis, A., Kafatos, F. C. and Maniatis, T. (1977). The primary structure of rabbit β -globin mRNA as determined from cloned DNA. *Cell* 10, 571-585.

Fraser, T. H. and Bruce, B. J. (1978). Chicken ovalbumin is synthesized and secreted by *Escherichia coli*. *Proc. Nat. Acad. Sci. USA* 75, 5936-5940.

Goeddel, D. V., Heyneker, H. L., Hozumi, T., Arentzen, R., Itakura, K., Yanewia, D. G., Ross, M. J., Miozzari, G., Crea, R. and Seeberg, P. H. (1979). Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone. *Nature* 281, 544-548.

Gray, H., Ostrander, D., Hodne, H. J., Legerski, R. and Roberson, D. (1975). Extracellular nucleases of *Pseudomonas* *Bal* 31. I. *Nucl. Acids Res.* 2, 1459-1492.

Hofstetter, H., Schamböck, A., Van Der Berg, J. and Weissmann, C. (1976). Specific excision of the inserted DNA segment from hybrid plasmids constructed by the poly(dA)-poly(dT) method. *Biochim. Biophys. Acta* 454, 587-591.

Jackson, D., Symons, R. and Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing λ phage genes and the galactose operon of *E. coli*. *Proc. Nat. Acad. Sci. USA* 67, 2904-2909.

Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Maniatis, T., Sim, G. K., Efstratiadis, A. and Kafatos, F. C. (1976). Amplification and characterization of a β -globin gene synthesized in vitro. *Cell* 8, 163-182.

Maurer, R., Meyer, B. and Ptashne, M. (1980). Gene regulation at the right operator of phage λ . I. *J. Mol. Biol.*, in press.

Maxam, A. and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Nat. Acad. Sci. USA* 74, 560-564.

Mercereau-Pujalon, O., Royal, A., Cami, B., Garapin, A., Krut, A., Gannon, F. and Kourilsky, P. (1978). Synthesis of an ovalbumin-like protein by *Escherichia coli* K12 harboring a recombinant plasmid. *Nature* 275, 505-510.

Miller, J. H. (1972). *Experiments in Molecular Genetics* (New York: Cold Spring Harbor Laboratory), p. 431.

Miller, J. (1978). The *lacI* gene: its role in *lac* operon control and its use as a genetic system. In *The Operon*, J. Miller and W. Reznikoff, eds. (New York: Cold Spring Harbor Laboratory), pp. 31-88.

Müller-Hill, B. and Kania, J. (1974). *Lac* repressor can be fused to β -galactosidase. *Nature* 249, 561-563.

Müller-Hill, B., Crapo, L. and Gilbert, W. (1968). Mutants that make more *lac* repressor. *Proc. Nat. Acad. Sci. USA* 59, 1259-1263.

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M. and Sauer, R. T. (1980). How the λ repressor and *cro* work. *Cell* 19, 1-11.

Roberts, T. and Lauer, G. (1979). Maximizing gene expression on a plasmid using recombination in vitro. *Methods in Enzymology*, 68 (New York: Academic Press), pp. 473-482.

Maximizing Cloned Gene Expression
553

- Roberts, T. M., Kacich, R. and Ptashne, M. (1979a). A general method for maximizing the expression of a cloned gene. *Proc. Nat. Acad. Sci. USA* 76, 760-764.
- Roberts, T. M., Bikel, I., Yocum, R. R., Livingston, D. M. and Ptashne, M. (1979b). Synthesis of simian virus 40 t antigen in *Escherichia coli*. *Proc. Nat. Acad. Sci. USA* 76, 5596-5600.
- Sancar, A., Hack, A. and Rupp, D. (1979). Simple method for identification of plasmid-coded proteins. *J. Bacteriol.* 137, 692-693.
- Sauer, R. (1978). DNA sequence of the bacteriophage λ *cl* gene. *Nature* 276, 301-302.
- Seeberg, P. H., Shine, J., Markal, J. A., Ivarice, R. D., Morris, J. A., Ulbrich, A., Baxter, J. D. and Goodman, H. M. (1978). Synthesis of growth hormone by bacteria. *Nature* 276, 795-798.
- Sheller, R. H., Dickerson, R. E., Boyer, H. W., Riggs, H. D. and Itakura, K. (1977). Chemical synthesis of restriction enzyme recognition sites useful for cloning. *Science* 196, 177-180.
- Shine, J. and Dalgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. *Nature* 254, 34-38.
- Steitz, J. A. (1979). Genetic signals and nucleotide sequences in messenger RNA. In *Biological Regulation and Control*, R. Golderger, ed. (New York: Plenum Press).
- Taniguchi, T. and Weissmann, C. (1978). Site-directed mutations in the indicator region of the bacteriophage Q β coat cistron and their effect on ribosome binding. *J. Mol. Biol.* 118, 533-565.
- Villa-Komaroff, L., Efstratiadis, A., Bromme, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. and Gilbert, W. (1978). A bacterial clone synthesizing proinsulin. *Proc. Nat. Acad. Sci. USA* 75, 3727-3731.

TAB RR



07/374311
S.N. 291,692
P/11/81

Case Docket 100/75

S-6654-CIP

521
MICROBIAL PRODUCTION
OF MATURE HUMAN
FIBROBLAST INTERFERON

CROSS-REFERENCE TO
RELATED APPLICATION

This is a continuation-in-part of our application of same title:
U.S. Serial No. 190799, filed September 25, 1980.

FIELD OF THE INVENTION

This invention relates to the microbial production, via recombinant DNA technology, of human fibroblast interferon for use in the treatment of viral and neoplastic diseases, and to the means and end products of such production.

0083L

-1-

BACKGROUND OF THE INVENTION

The publications and other materials referred to herein to illuminate the background of the invention and, in particular cases, to provide additional detail respecting its practice are incorporated herein by reference and, for convenience; are numerically referenced in the following text and respectively grouped in the appended bibliography.

Recombinant DNA Technology

With the advent of recombinant DNA technology, the controlled microbial production of an enormous variety of useful polypeptides has become possible. Already in hand are bacteria modified by this technology to permit the production of such polypeptide products such as somatostatin, the (component) A and B chains of human insulin, human growth hormone. More recently, recombinant DNA techniques have been used to occasion the bacterial production of proinsulin, thymosin alpha 1, (an immune potentiating substance produced by the thymus) and leukocyte interferon.

The workhorse of recombinant DNA technology is the plasmid, a non-chromosomal loop of double-stranded DNA found in bacteria and other microbes, oftentimes in multiple copies per cell. Included in the information encoded in the plasmid DNA is that required to reproduce the plasmid in daughter cells (i.e., a "replicon") and ordinarily, one or more selection characteristics such as, in the case of bacteria, resistance to antibiotics which permit clones of the host cell containing the plasmid of interest to be recognized and preferentially grown in selective media. The utility of plasmids lies in the fact that they can be specifically cleaved by one or another restriction endonuclease or "restriction enzyme", each of which recognizes a different site on the plasmidic DNA. Thereafter heterologous genes or gene fragments may be inserted into the plasmid by endwise joining at the cleavage site or at reconstructed ends adjacent to the cleavage site. DNA recombination is performed outside the cell, but the resulting "recombinant" plasmid can be introduced into it by a process known as transformation and large quantities of the heterologous gene-containing recombinant plasmid obtained by growing the transformant. Moreover, where the gene is properly inserted with reference to portions of the plasmid which govern the transcription and translation of the encoded DNA message, the resulting expression vehicle can be used to actually produce the polypeptide sequence for which the inserted gene codes, a process referred to as expression.

Expression is initiated in a region known as the promoter which is recognized by and bound by RNA polymerase. In some cases, as in the tryptophan or "trp" promoter preferred in the practice of the present invention, promoter regions are overlapped by "operator" regions to form a combined promoter-operator. Operators are DNA sequences which are recognized by so-called repressor proteins which serve to regulate the frequency of transcription initiation at a particular promoter. The polymerase travels along the DNA, transcribing the information contained in the coding strand from its 5' to 3' end into messenger RNA which is in turn translated into a polypeptide having the amino acid sequence for which the DNA codes. Each amino acid is encoded by a nucleotide triplet or "codon" within what may for present purposes be referred to as the "structural gene", i.e. that part which encodes the amino acid sequence of the expressed product. After binding to the promoter, the RNA polymerase first transcribes nucleotides encoding a ribosome binding site, then a translation initiation or "start" signal (ordinarily ATG, which in the resulting messenger RNA becomes AUG), then the nucleotide codons within the structural gene itself. So-called stop codons are transcribed at the end of the structural gene whereafter the polymerase may form an additional sequence of messenger RNA which, because of the presence of the stop signal, will remain untranslated by the ribosomes. Ribosomes bind to the binding site provided on the messenger RNA, in bacteria ordinarily as the mRNA is being formed, and themselves produce the encoded polypeptide, beginning at the translation start signal and ending at the previously mentioned stop signal. The desired product is produced if the sequences encoding the ribosome binding site are positioned properly with respect to the AUG initiator codon and if all remaining codons follow the initiator codon in phase. The resulting product may be obtained by lysing the host cell and recovering the product by appropriate purification from other bacterial protein.

Fibroblast Interferon

Human fibroblast interferon (FIF) is an antiviral protein which also exhibits a wide range of other biological activities (see ref. 1 for review). It has reportedly been purified to homogeneity as a single polypeptide of 19,000-20,000 molecular weight having a specific activity of 2 to 10×10^8 units/mg (2,3). The sequence of the 13 NH_2 -terminal amino acids of FIF has been determined (4). Houghton *et al.* (5) have used synthetic deoxyoligonucleotides (predicted from this amino acid sequence) to determine the sequence of the 276 5'-terminal nucleotides of FIF mRNA. Taniguchi *et al.* (6) and Derynck *et al.* (7) have recently

employed RNA selection procedures to identify cloned cDNA copies of FIF mRNA in *E. coli*. See also Taniguchi et al., *Gene* 10, 11 (1980) and *Proc. Natl. Acad. Sci. (U.S.A.)* 77, 5230 (1980) and *Nature* 285, 547 (1980).

While isolation from donor fibroblasts has provided sufficient material for partial characterization and limited clinical studies with homogeneous fibroblast interferon, it is a totally inadequate source for the amounts of interferon needed for large scale clinical trials and for broad scale prophylactic and/or therapeutic use thereafter. Indeed, presently clinical investigations employing human fibroblast-derived interferon in antitumor and antiviral testing have principally been confined to crude (< 1 percent pure) preparations of the material, and long lead times for the manufacture of sufficient quantities, even at unrealistic price levels, have critically delayed investigation on an expanded front.

We perceived that application of recombinant DNA technology would be the most effective way of providing large quantities of fibroblast interferon which, despite the absence in material so produced of the glycosylation characteristic of human-derived material, could be employed clinically in the treatment of a wide range of viral and neoplastic diseases.

More particularly, we proposed and have since succeeded in producing mature human fibroblast interferon microbially, by constructing a gene therefor which could then be inserted in microbial expression vehicles and expressed under the control of microbial gene regulatory controls.

Our approach to obtaining a fibroblast gene involved the following tasks:

1. Partial amino acid sequences would be obtained by characterization of fibroblast interferon purified to essential homogeneity, and sets of synthetic DNA probes constructed whose codons would, in the aggregate, represent all the possible combinations capable of encoding the partial amino acid sequences.

2. Bacterial colony banks would be prepared containing cDNA from induced messenger RNA. The probes of part (1) would be used to prime the synthesis of radio-labelled single stranded cDNA for use as hybridization probes. The synthetic probes would hybridize with induced mRNA as template and be extended by reverse transcription to form induced,

radio-labelled cDNA. Clones from the colony bank that hybridized to radio-labelled cDNA obtained in this manner would be investigated further to confirm the presence of a full-length interferon encoding gene. Any partial length putative gene fragment obtained would itself be used as a probe for the full-length gene.

3. The full-length gene obtained above would be tailored, using synthetic DNA, to eliminate any leader sequence that might prevent microbial expression of the mature polypeptide and to permit appropriate positioning in an expression vehicle relative to start signals and the ribosome binding site of a microbial promoter. Expressed interferon would be purified to a point permitting confirmation of its character and determination of its activity notwithstanding the absence of glycosylation.

SUMMARY OF INVENTION

C
C
C
D
A series of replicable plasmidic expression vehicles have been constructed which direct the high level synthesis in transformant microorganisms of a mature polypeptide with the properties of authentic human fibroblast interferon. The product polypeptide exhibits the amino acid sequence of such interferon and is active in in vitro testing despite the lack of glycosylation characteristic of the human-derived material. Reference herein to the expression of "mature fibroblast interferon," connotes the bacterial or other microbial production of an interferon molecule unaccompanied by associated glycosylation and the presequence that immediately attends mRNA translation of the human fibroblast interferon genome. Mature fibroblast interferon, according to the present invention, is immediately expressed from a translation start signal (ATG) which also encodes the first amino acid codon of the natural product. The presence or absence of the methionine first amino acid in the microbially expressed product is governed by a kinetic phenomenon dependent on fermentation growth conditions and/or levels of expression in the transformant host. Mature fibroblast interferon could be expressed together with a conjugated protein other than the conventional leader, the conjugate being specifically cleavable in an intra- or extracellular environment. See British Patent Publication No. 2007676A. Finally, the mature interferon could be produced in conjunction with a microbial "signal" peptide which transports the conjugate to the cell wall, where the signal is processed away and the mature polypeptide secreted.

Figures 1 to 5 appended hereto are described in the detailed text infra. Figure 6 schematically depicts the construction of plasmids coding for the direct expression of mature fibroblast interferon. Restriction sites and residues are as shown ("Pst I", etc.). "Ap^R" and "Tc^R" connote portions of the plasmid(s) which express, respectively, ampicillin and tetracycline resistance. The legend "p o" is an abbreviation for "promoter operator."

DESCRIPTION OF THE PREFERRED EMBODIMENTS

A. Microorganisms Employed

The work described involved use of the microorganism: E. coli K-12 strain 294 (end A, thi⁻, hsr⁻, hsm⁺), as described in British Patent Publication No. 2055382 A. ~~This strain has been deposited with the American Type Culture Collection, ATCC Accession No. 31446. All~~ recombinant DNA work was performed in compliance with applicable guidelines of the National Institutes of Health.

The invention, in its most preferred embodiments, is described with reference to E. coli, including not only strain E. coli K-12 strain 294, defined above, but also other known E. coli strains such as E. coli B, E. coli x 1776 and E. coli W 3110, or other microbial strains many of which are deposited and (potentially) available from recognized microorganism depository institutions, such as the American Type Culture Collection (ATCC)--cf. the ATCC catalogue listing. See also German Offenlegungsschrift 2644432. These other microorganisms include, for example, Bacilli such as *Bacillus subtilis* and other enterobacteriaceae among which can be mentioned as examples *Salmonella typhimurium* and *Serratia marcescens*, utilizing plasmids that can replicate and express heterologous gene sequences therein. Yeast, such as *Saccharomyces cerevisiae*, may also be employed to advantage as host organism in the preparation of the interferon proteins hereof by expression of genes coding therefor under the control of a yeast promoter. (See the copending U.S. patent application of Hitzeman et al., filed February 25, 1981 (Attorney Docket No. 100/43), assignee Genentech, Inc. et al., which is incorporated herein by reference.)

MATERIALS AND METHODS

General methods.

Restriction enzymes were purchased from New England Biolabs and used

as directed. Plasmid DNA was prepared by a standard cleared lysate procedure (8) and purified by column chromatography on Biogel A-5DM (Bio-Rad). DNA sequencing was performed using the method of Maxam and Gilbert (9). DNA restriction fragments were isolated from polyacrylamide gels by electroelution. DNA fragments were radiolabeled for use as hybridization probes by the random calf thymus DNA priming procedure of Taylor et al. (10). In situ colony hybridizations were performed by the Grunstein-Hogness procedure (11).

Chemical synthesis of deoxyoligonucleotides.

The deoxyoligonucleotides were synthesized by the modified phosphotriester method in solution (12), using trideoxynucleotides as building blocks (13). The materials and general procedures were similar to those described (14). The six pools of primers (Fib 1-6) containing four dodecanucleotides each were obtained by separately coupling two hexamer pools (of two different 5'-terminal sequences each) with three different hexamer pools (of two different 3'-terminal sequences each).

Induction of fibroblasts.

Human fibroblasts (cell line GM-2504A) were grown as described previously (15). Growth medium (Eagle's minimal essential medium containing 10 percent fetal calf serum) was removed from roller bottles (Corning, 850 cm²) and replaced with 50 ml growth medium containing 50 µg/ml of poly(I):poly(C) (PL Biochemicals) and 10 µg/ml cycloheximide. This induction medium was removed after 4 hours at 37°C and cell monolayers were washed with PBS (0.14M NaCl, 3mM KCl, 1.5 mM KH₂PO₄, 8mM Na₂HPO₄). Each bottle was incubated at 37°C with 10 ml of a trypsin - EDTA solution (Gibco 610-5305) until cells were detached, and fetal calf serum was added to a concentration of 10 percent. Cells were spun for 15 minutes at 500 x g and pellets were resuspended in PBS, pooled, and resedimented. Cells were frozen in liquid nitrogen. Approximately 0.17g of cells were obtained per roller bottle.

Preparation and assay of interferon mRNA.

Poly(A)-containing mRNA was prepared from human fibroblasts by phenol extraction and oligo(dT)-cellulose chromatography as described elsewhere (16). The poly (A) containing RNA was enriched for interferon mRNA by centrifugation on a linear 5 percent to 20 percent (w/v) sucrose gradient. The RNA samples were heated to 80°C for 2 minutes, rapidly

cooled; layered over the gradient, and centrifuged for 20 hours at 30,000 rpm at 4°C in a Beckman SW-40 rotor. Fractions were collected, ethanol precipitated, and dissolved in H₂O.

One microgram samples of mRNA were injected into Xenopus laevis oocytes as described previously (17,18). The injected oocytes were incubated 24 hours at 21°C, homogenized, and centrifuged for 5 minutes at 10,000 x g. The interferon in the supernatant was determined by the cytopathic effect (CPE) inhibition assay (1) using Sindbis virus and human diploid (WISH) cells. Interferon titers of 1,000 to 6,000 units recovered (NIH reference standard) per microgram of RNA injected were routinely obtained for the 12S species of mRNA.

Synthesis and cloning of cDNA.

Single stranded cDNA was prepared in 100 µl reactions containing 5 µg of 12S fraction mRNA, 20 mM Tris-HCl (pH 8.3), 20 mM KCl, 8mM MgCl₂, 30 mM β-mercaptoethanol, 100 µCi of (α³²P)dCTP (Amersham) and 1mM dATP, dCTP, dGTP, dTTP. The primer was the synthetic Hind III decamer dCCAAGCTTGG (19), which had been extended at the 3' terminus with about 20 to 30 deoxythymidine residues using terminal deoxynucleotidyl transferase (20). 100 units of AMV reverse transcriptase were added and the reaction mixture was incubated at 42°C for 30 minutes. The second strand DNA synthesis was carried out as described previously (21). The double stranded cDNA was treated with 1200 units of S1 nuclease (Miles Laboratories) for 2 hours at 37°C in 25 mM sodium acetate (pH 4.5), 1mM ZnCl₂, 0.3M NaCl. After phenol extraction the mixture was separated electrophoretically on an 8 percent polyacrylamide gel. cDNA (~ 0.5' µg) ranging from 550 to 1500 base pairs in size was recovered by electroelution. A 20 ng aliquot was extended with deoxyC residues using terminal deoxynucleotidyl transferase (20), and annealed with 100 ng of pBR322 which had been cleaved with Pst I and tailed with deoxyG residues (20). The annealed mixture was used to transform E. coli K-12 strain 294 (22) by a published procedure (23). Strain 294 was used throughout in the work described here, and has been deposited with the American Type Culture collection, accession no. 31446.

Preparation of induced and uninduced ³²P-cDNA probes. 5 µg of 12S mRNA were combined with either 2 µg of oligo (dT)₁₂₋₁₈ (Collaborative Research) or 5 µg of each synthetic primer pool (Fib 1 to Fib 6) in 60 µl of 10mM Tris-HCl (pH 8), 1 mM EDTA. The mixtures were boiled 3 minutes,

and quenched on ice. 60 μ l of 40 mM Tris-HCl (pH 8.3), 40 mM KCl, 16mM $MgCl_2$, 60 mM α -mercaptoethanol, 1 mM dATP, dGTP, dTTP and 5×10^{-7} M (α - ^{32}P) dCTP (Amersham, 2,000 - 3,000 Ci/mmole) was added to each template-primer mix at 0°C. After the addition of 100 units of AMV reverse transcriptase, the reactions were incubated at 42°C for 30 minutes and purified by passage over 10 ml Sephadex G-50-columns. The products were treated with 0.3N NaOH for 30 minutes at 70°C, neutralized, and ethanol precipitated.

The ^{32}P -cDNAs were combined with 100 μ g of poly(A) mRNA from uninduced fibroblasts in 50 μ l of 0.4M sodium phosphate (pH 6.8), 0.1 percent SDS. The mixtures were heated at 98°C for 5 minutes and allowed to anneal 15 hours at 45°C. The DNA-RNA hybrids (containing uninduced cDNA sequences) were separated from single-stranded DNA (induced cDNA sequences) by chromatography on hydroxyapatite as described by Galau *et al.* (24). The DNA-RNA hybrids were treated with alkali to remove RNA.

Screening of recombinant plasmids with ^{32}P -cDNA probes.

Approximately 1 μ g samples of plasmid DNA were prepared from individual transformants by a published procedure (25). The DNA samples were linearized by digestion with Eco RI, denatured in alkali, and applied to each of three nitrocellulose filters (Schleicher and Schuell, BA85) by the dot hybridization procedure (26). The filters were hybridized with the ^{32}P -cDNA probes for 16 hours at 42°C in 50 percent formamide, 10x Denhardt's solution (27), 6xSSC, 40 mM Tris-HCl (pH 7.5), 2mM EDTA, 40 μ g/ml yeast RNA. Filters were washed with 0.1xSSC, 0.1 percent SDS twice for 30' at 42°C, dried, and exposed to Kodak XR-2 x-ray film using Dupont Lightning-Plus intensifying screens at -80°C.

Construction of plasmids for direct expression of FIF.

The synthetic primers I (dATGAGCTACAAC) and II (dCATGAGCTACAAC) were phosphorylated using T4 polynucleotide kinase and (γ - ^{32}P)ATP (Amersham) to a specific activity of 700 Ci/mmole as described previously (28). Primer repair reactions were performed as follows: 250 pmoles of the ^{32}P -primers were combined with 8 μ g (10 pmole) of a 1200 bp Hha I restriction fragment containing the FIF cDNA sequence. The mixture was ethanol precipitated, resuspended in 50 μ l H_2O , boiled 3 minutes, quenched in a dry ice-ethanol bath, and combined with a 50 μ l solution of 20mM Tris-HCl (pH 7.5), 14 mM $MgCl_2$, 120 mM NaCl, 0.5 mM dATP, dCTP, dGTP, dTTP at 0°C. 10 units of DNA polymerase I Klenow fragment (Boehringer-Mannheim) were added and the mixture was incubated at 37°C

for 4 1/2 hours. Following extraction with phenol/CHCl₃ and restriction with Pst I, the desired product was purified on a 6 percent polyacrylamide gel. Subsequent ligations were done at room temperature (cohesive termini) or 4°C (blunt ends) using previously detailed conditions (21,28).

Assay for interferon expression in E. coli.

Bacterial extracts were prepared for IF assay as follows: One ml cultures were grown overnight in LB (29) containing 5 µg/ml tetracycline, then diluted into 25 ml of M9 medium (29) containing 0.2 percent glucose, 0.5 percent casamino acids and 5 µg/ml tetracycline. 10 ml samples were harvested by centrifugation when A₅₅₀ (Absorbance at 550 nanometers) reached 1.0. The cell pellets were quickly frozen in a dry ice-ethanol bath and cleared lysates were prepared as described by Clewell (8). Interferon activity in the supernatants was determined by comparison with NIH FIF standards using cytopathic effect (CPE) inhibition assays as reviewed previously (1). Two different assays were used: (a) WISH (human amnion) cells were seeded in microtiter dishes. Samples were added 16 to 20 hours later and diluted by serial 2-fold dilution. Sindbis virus was added after at least 3 hours of incubation. Plates were stained 20 to 24 hours later with crystal violet. (b) MDBK (bovine kidney) cell line was seeded simultaneously with 2-fold dilutions of samples. Vesicular stomatitis virus was added after 2 to 3 hours incubation and plates were stained with crystal violet 16 to 18 hours later. To test pH 2 stability bacterial extracts and standards were diluted in minimal essential medium to a concentration of 1000 units/ml. One ml aliquots were adjusted to pH 2 with 1N HCl, incubated at 4°C for 16 hours, and neutralized by addition of NaOH. IF activity was determined by the CPE inhibition assay using human amnion cells. To establish antigenic identity 25 µl aliquots of the 1000 U/ml interferon samples (untreated) were incubated with 25 µl of rabbit antihuman leukocyte interferon for 60' at 37°C, centrifuged at 12,000 x g for 5 minutes and the supernatant assayed. Fibroblast and leukocyte interferon standards were obtained from the National Institutes of Health. Rabbit antihuman leukocyte interferon was obtained from the National Institute of Allergy and Infectious Diseases.

RESULTS

Chemical synthesis of primer pools complementary to FIF mRNA.

The amino-terminal protein sequence of human fibroblast interferon

(4) permitted us to deduce the 24 possible mRNA sequences which could code for the first four amino acids. The 24 complementary deoxyoligonucleotides were synthesized in 6 pools of 4 dodecamers each (Figure 1).

The six pools of 4 deoxyoligonucleotides each were synthesized by a modified phosphotriester method that has been used previously for the rapid synthesis of oligonucleotides in solution (12) and on solid phase (14). The basic strategy involved reacting two different 3'-blocked trimers with an excess of a single 5'-protected trimer to yield a pool of two hexamers, each represented equally. The coupling of two pools, each containing two hexamers, then resulted in a pool of four dodecamers.

Identification of FIF cDNA clones.

Using 12S mRNA from induced human fibroblasts (1,000 units IF activity per μg in oocyte assay), double stranded cDNA was prepared and inserted into pBR322 at the Pst I site by the standard dG:dC tailing method (20). A fibroblast cDNA library consisting of 30,000 ampicillin-sensitive, tetracycline-resistant transformants of *E. coli* K-12 strain 294 was obtained from 2D ng of cDNA ranging in size from 550 to 1300 base pairs. Plasmid DNA was prepared from 60D of the transformants and applied to 3 sets of nitrocellulose filters as described in Materials and Methods.

The approach followed in the identification of hybrid plasmids containing fibroblast interferon cDNA sequences was similar to that used to identify human leukocyte interferon recombinant plasmids (30). Radiolabeled cDNA hybridization probes were prepared using either the 24 synthetic dodecamers or oligo(dT)₁₂₋₁₈ as primers and 12S RNA from induced fibroblasts (5000 units/ μg in oocytes) as template. The ³²P-cDNAs (specific activity $>5 \times 10^8$ cpm/ μg) obtained were hybridized to a large excess of mRNA isolated from uninduced human fibroblasts, and the mRNA-cDNA hybrids were separated from unreacted cDNA by hydroxyapatite chromatography (24). The single stranded cDNA fractions should be enriched for sequences which are present in induced fibroblasts but absent in uninduced cells, and the mRNA-cDNA hybrids should represent sequences common to both induced and uninduced cells. Approximately 4×10^6 cpm of single stranded cDNA (hybridization probe A) and 8×10^6 cpm of cDNA-mRNA hybrids were obtained using oligo(dT)₁₂₋₁₈ primed cDNA; 1.5×10^6 cpm of single stranded (hybridization probe B) and 1.5×10^5 cpm of hybrids were obtained from cDNA primed using synthetic dodecamer pools Fib 1-6. The cDNA-mRNA hybrids from both fractionations were combined, the RNA hydrolyzed by treatment with

alkali, and the ^{32}P -cDNA used as hybridization probe C. Many of the 600 plasmid samples hybridized with both probes A and C, indicating that the hybridization reactions between uninduced mRNA and ^{32}P -cDNA (prior to the hydroxyapatite fractionation step) had not gone to completion. However, only one of the 600 plasmids (pF526) hybridized strongly with the specifically primed, induced cDNA probe B (Figure 2). Plasmid pF526 also hybridized with the total oligo(dT)₁₂₋₁₈ primed, induced cDNA probe A, and failed to give detectable hybridization to the combined uninduced probe C.

Pst I digestion of pF526 showed the cloned cDNA insert to be about 550 base pairs long, probably too short to contain the entire coding region for a protein the size of fibroblast interferon. Therefore, a ^{32}P -labeled DNA probe was prepared from this Pst I fragment by random priming with calf thymus DNA (10). This probe was used to screen 2000 individual colonies from a newly constructed fibroblast cDNA library (the new cDNA library was prepared using 12S mRNA from induced fibroblasts having a titer of 6,000 units/ml in the oocyte assay system). Sixteen clones hybridized to the probe. Plasmids prepared from the majority of these released two fragments when cleaved with Pst I, indicating that the cDNA contained an internal Pst I site. Clone pFIF3 contained the largest cDNA insert, about 800 base pairs. The DNA sequence of the insert was determined by the Maxam-Gilbert procedure (9) and is shown in Figure 3. The amino acid sequence of human fibroblast interferon predicted from the nucleotide sequence is identical to that reported recently by Taniguchi *et al.* (31) and by Derynck *et al.* (7) from DNA sequencing of FIF cDNA clones. A precursor or signal peptide of 21 amino acids is followed by a mature interferon polypeptide of 166 amino acids, a stretch of 196 3'-untranslated nucleotides and a poly(A) tail. The NH₂-terminal 20 amino acids of mature FIF have been directly determined by protein microsequencing and are the same as those predicted from the DNA sequence.

Direct expression of fibroblast interferon.

To express high levels of mature fibroblast interferon in *E. coli*, initiation of protein synthesis must occur at the ATG codon of the mature polypeptide (amino acid 1) rather than at the ATG of the signal peptide (amino acid 51) (Figure 3).

Our approach to removing the signal peptide coding regions from pFIF3 is depicted in Figure 4. A 1200 bp DNA fragment which contained the entire cDNA insert was isolated from a polyacrylamide gel after digesting

pFIF3 with Hha I. Two separate synthetic deoxyoligonucleotide primers, dATGAGCTACAAC(I) and dCATGAGCTACAAC(II), were prepared. Both primers contain the coding sequence for the first four amino acids of mature fibroblast interferon; primer II has an additional 'C' at the 5'-terminus. Primer repair reactions and subsequent ligations were carried out separately for primers I and II, and gave nearly identical results. Therefore, only reactions using primer I are discussed in detail here. The primers were 5'-radiolabeled using (γ - 32 P)ATP and T4 polynucleotide kinase, combined with the 1200 bp Hha I DNA fragment and the mixture denatured by boiling. Following hybridization of the primer to the denatured Hha I DNA fragment, E. coli DNA polymerase I Klenow fragment (33) was used to catalyze the repair synthesis of the plus (top) strand (Figure 4). In addition, the associated 3' \rightarrow 5' exonuclease activity of the Klenow fragment removed the 3'-protruding end from the minus (bottom) strand, leaving a flush end. Analysis of samples of the reaction mixture by polyacrylamide gel electrophoresis indicated that the repair synthesis did not go to completion, but stopped at several discrete sites. Therefore, the entire reaction mixture was treated with Pst I and the desired 141 bp fragment (180,000 Cerenkov cpm; \sim 0.3 pmole) was purified by polyacrylamide gel electrophoresis (Figure 5). Ligation of this fragment to 1 μ g (\sim 4 pmole) of the 363 bp Pst I-Bgl II fragment isolated from pFIF3 (Fig. 4), followed by Bgl II digestion, yielded 50,000 Cerenkov cpm (\sim 0.1 pmole, \sim 30 ng) of the 504 bp DNA fragment containing the entire coding sequence for mature fibroblast interferon. The same reactions using primer II gave 83,000 cpm (\sim 0.15 pmole, \sim 50 ng) of 505 bp product.

The construction of plasmids which direct the synthesis of human fibroblast interferon is outlined in Figure 6. Separate expression plasmids were constructed which placed FIF synthesis under the control of the E. coli lac or trp promoter-operator systems. Both of these systems have proven useful for the direct expression of eukaryotic genes in E. coli: human growth hormone has been efficiently synthesized using the lac system (21) and human leukocyte interferon has been produced at high levels using the trp system (30) and Nature 287, 411 (1980).

pBKH trp was digested with EcoRI restriction enzyme and the resulting fragment isolated by PAGE and electroelution. EcoRI-digested plasmid pSom 11 (K. Itakura et al., Science 198, 1056 (1977); G.B. patent publication no. 2 007 676 A) was combined with the above fragment. The mixture was ligated with T₄ DNA ligase as previously described and the resulting DNA transformed into E. coli K-12 strain 294 as previously.

described. Transformant bacteria were selected on ampicillin-containing plates. Resulting ampicillin-resistant colonies were screened by colony hybridization (M. Gruenstein *et al.*, Proc Nat'l Acad Sci USA 72, 3951-3965 [1975]) using as a probe the trp promoter-operator-containing the above fragment isolated from pBRHtrp, which had been radioactively labelled with p³². Several colonies shown positive by colony hybridization were selected, plasmid DNA was isolated and the orientation of the inserted fragments determined by restriction analysis employing restriction enzymes BglII and BamHI in double-digestion. *E. coli* 294 containing the plasmid designated pSOM7Δ2, which has the trp promoter-operator fragment in the desired orientation was grown in LB medium containing 10 µg/ml ampicillin. The cells were grown to optical density 1 (at 550 mμ), collected by centrifugation and resuspended in M9 media in tenfold dilution. Cells were grown for 2-3 hours, again to optical density 1, then lysed and total cellular protein analyzed by SDS (sodium dodecyl sulfate) urea (15 percent) polyacrylamide gel electrophoresis (J.V. Maizel Jr. *et al.*, Meth Viral 5, 180-246 [1971]).

Plasmid pBR322 was Hind III digested and the protruding Hind III ends in turn digested with SI nuclease. The SI nuclease digestion involved treatment of 10 µg of Hind III-cleaved pBR322 in 30 µl SI buffer (0.3 M NaCl, 1 mM ZnCl₂, 25 mM sodium acetate, pH 4.5) with 300 units SI nuclease for 30 minutes at 15°C. The reaction was stopped by the addition of 1 µl of 30 X SI nuclease stop solution (0.8M tris base, 50 mM EDTA). The mixture was phenol extracted; chloroform extracted and ethanol precipitated, then EcoRI digested as previously described and the large fragment (1) obtained by PAGE procedure followed by electroelution. The fragment obtained has a first EcoRI sticky end and a second, blunt end whose coding strand begins with the nucleotide thymidine.

Plasmid pSom7 Δ2, as prepared above, was Bgl II digested and the Bgl II sticky ends resulting made double stranded with the Klenow polymerase I procedure using all four deoxynucleotide triphosphates. EcoRI cleavage of the resulting product followed by PAGE and electroelution of the small fragment (2) yielded a linear piece of DNA containing the tryptophan promoter-operator and codons of the LE' "proximal" sequence upstream from the Bgl II site ("LE'(p)"). The product had an EcoRI end and a blunt end resulting from filling in the Bgl II site. However, the Bgl II site is reconstituted by ligation of the blunt end of the above fragment (2) to the blunt end of the above prepared fragment (1). Thus, the two fragments were ligated in the presence of T₄ DNA ligase to form the

recirculated plasmid pHKY 10 which was propagated by transformation into competent E. coli strain 294 cells.

Plasmid pGM1 carries the E. coli tryptophan operon containing the deletion Δ LE1413 (G.F. Miozzari, et al., (1978) J. Bacteriology 133, 1457-1466)) and hence expresses a fusion protein comprising the first 6 amino acids of the trp leader and approximately the last third of the trp E polypeptide (hereinafter referred to in conjunction as LE'), as well as the trp D polypeptide in its entirety, all under the control of the trp promoter-operator system. The plasmid, 20 μ g, was digested with the restriction enzyme PvuII which cleaves the plasmid at five sites. The gene fragments were next combined with EcoRI linkers (consisting of a self complementary oligonucleotide of the sequence: pCATGAATTCATG) providing an EcoRI cleavage site for a later cloning into a plasmid containing an EcoRI site. The 20 μ g of DNA fragments obtained from pGM1 were treated with 10 units T_4 DNA ligase in the presence of 200 pico moles of the 5'-phosphorylated synthetic oligonucleotide pCATGAATTCATG and in 20 μ l T_4 DNA ligase buffer (20mM tris, pH 7.6, 0.5 mM ATP, 10 mM $MgCl_2$, 5 mM dithiothreitol) at 4°C overnight. The solution was then heated 10 minutes at 70°C to halt ligation. The linkers were cleaved by EcoRI digestion and the fragments, now with EcoRI ends were separated using 5 percent polyacrylamide gel electrophoresis (hereinafter "PAGE") and the three largest fragments isolated from the gel by first staining with ethidium bromide, locating the fragments with ultraviolet light, and cutting from the gel the portions of interest. Each gel fragment, with 300 microliters 0.1xTBE, was placed in a dialysis bag and subjected to electrophoresis at 100 v for one hour in 0.1xTBE buffer (TBE buffer contains: 10.8 gm tris base, 5.5 gm boric acid, 0.09 gm Na_2EDTA in 1 liter H_2O). The aqueous solution was collected from the dialysis bag, phenol extracted, chloroform extracted and made 0.2 M sodium chloride, and the DNA recovered in water after ethanol precipitation. The trp promoter-operator-containing gene with EcoRI sticky ends was identified in the procedure next described, which entails the insertion of fragments into a tetracycline sensitive plasmid which, upon promoter-operator insertion, becomes tetracycline resistant.

Plasmid pBRH1 (R.I. Rodriguez, et al., Nucleic Acids Research 6, 3267-3287 [1979]) expresses ampicillin resistance and contains the gene for tetracycline resistance but, there being no associated promoter, does not express that resistance. The plasmid is accordingly tetracycline sensitive. By introducing a promoter-operator system in the EcoRI site, the plasmid can be made tetracycline resistant.

pBRH1 was digested with EcoRI and the enzyme removed by phenol extraction followed by chloroform extraction and recovered in water after ethanol precipitation. The resulting DNA molecule was, in separate reaction mixtures, combined with each of the three DNA fragments obtained above and ligated with T₄ DNA ligase as previously described. The DNA present in the reaction mixture was used to transform competent *E. coli* K-12 strain 294, K. Backman *et al.*, Proc Nat'l Acad Sci USA 73, 4174-4198. [1976]) by standard techniques (V. Hershfild *et al.*, Proc Nat'l Acad Sci USA 71, 3455-3459 [1974]) and the bacteria plated on LB plates containing 20 µg/ml ampicillin and 5 µg/ml tetracycline. Several tetracycline-resistant colonies were selected, plasmid DNA isolated and the presence of the desired fragment confirmed by restriction enzyme analysis. The resulting plasmid is designated pBRHtrp.

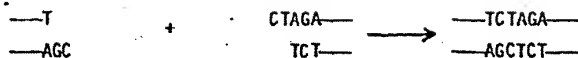
An EcoRI and BamHI digestion product of the viral genome of hepatitis B was obtained by conventional means and cloned into the EcoRI and BamHI sites of plasmid pGH6 (D.V. Goeddel *et al.*, Nature 281, 544 [1979]) to form the plasmid pHS32. Plasmid pHS32 was cleaved with XbaI, phenol extracted, chloroform extracted and ethanol precipitated. It was then treated with 1 µl *E. coli* polymerase I, Klenow fragment (Boehringer-Mannheim) in 30 µl polymerase buffer (50 mM potassium phosphate pH 7.4, 7mM MgCl₂, 1 mM β-mercaptoethanol) containing 0.1mM dTTP and 0.1mM dCTP for 30 minutes at 0°C then 2 hr. at 37°C. This treatment causes 2 of the 4 nucleotides complementary to the 5' protruding end of the XbaI cleavage site to be filled in:

| | | | |
|----|--------|----|--------|
| 5' | CTAGA— | 5' | CTAGA— |
| 3' | T— | 3' | TCT— |

Two nucleotides, dC and dT, were incorporated giving an end with two 5' protruding nucleotides. This linear residue of plasmid pHS32 (after phenol and chloroform extraction and recovery in water after ethanol precipitation) was cleaved with EcoRI. The large plasmid fragment was separated from the smaller EcoRI-XbaI fragment by PAGE and isolated after electroelution. This DNA fragment from pHS32 (0.2 µg), was ligated, under conditions similar to those described above, to the EcoRI-Taq I fragment of the tryptophan operon (~0.01 µg), derived from pBRHtrp.

In the process of ligating the fragment from pHS32 to the Eco RI-Taq I fragment, as described above, the Taq I protruding end is ligated to the XbaI remaining protruding end even though it is not completely

Watson-Crick base-paired:



A portion of this ligation reaction mixture was transformed into *E. coli*-294 cells, heat treated and plated on LB plates containing ampicillin. Twenty-four colonies were selected, grown in 3 ml LB media, and plasmid isolated. Six of these were found to have the *Xba*I site regenerated via *E. coli* catalyzed DNA repair and replication:



These plasmids were also found to cleave both with *Eco*RI and *Hpa*I and to give the expected restriction fragments. One plasmid, designated pTrp 14, was used for expression of heterologous polypeptides, as next discussed.

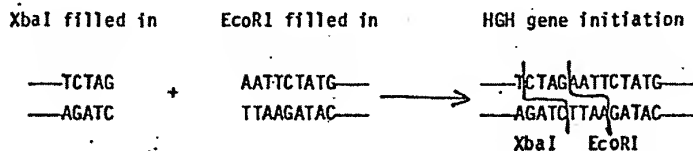
The plasmid pHGH 107 (D.V. Goeddel et al, *Nature*, 281, 544, 1979) contains a gene for human growth hormone made up of 23 amino acid codons produced from synthetic DNA fragments and 163 amino acid codons obtained from complementary DNA produced via reverse transcription of human growth hormone messenger RNA. This gene, though it lacks the codons of the "pre" sequence of human growth hormone, does contain an ATG translation initiation codon. The gene was isolated from 10 µg pHGH 107 after treatment with *Eco*RI followed by *E. coli* polymerase I Klenow fragment and dTTP and dATP as described above. Following phenol and chloroform extraction and ethanol precipitation the plasmid was treated with *Bam*HI.

The human growth hormone ("HGH") gene-containing fragment was isolated by PAGE followed by electroelution. The resulting DNA fragment also contains the first 350 nucleotides of the tetracycline-resistance structural gene, but lacks the tetracycline promoter-operator system so that, when subsequently cloned into an expression plasmid, plasmids containing the insert can be located by the restoration of tetracycline resistance. Because the *Eco*RI end of the fragment has been filled in by the Klenow polymerase I procedure, the fragment has one blunt and one sticky end, ensuring proper orientation when later inserted into an expression plasmid.

The expression plasmid pTrp14 was next prepared to receive the HGH

gene-containing fragment prepared above. Thus, pTrp14 was XbaI digested and the resulting sticky ends filled in with the Klenow polymerase I procedure employing dATP, dTTP, dGTP and dCTP. After phenol and chloroform extraction and ethanol precipitation the resulting DNA was treated with BamHI and the resulting large plasmid fragment isolated by PAGE and electroelution. The pTrp14-derived fragment had one blunt and one sticky end, permitting recombination in proper orientation with the HGH gene containing fragment previously described.

The HGH gene fragment and the pTrp14 Δ Xba-BamHI fragment were combined and ligated together under conditions similar to those described above. The filled in XbaI and EcoRI ends ligated together by blunt end ligation to recreate both the XbaI and the EcoRI site:



This construction also recreates the tetracycline resistance gene. Since the plasmid pHGH 107 expresses tetracycline resistance from a promoter lying upstream from the HGH gene (the lac promoter), this construction, designated pHGH 207, permits expression of the gene for tetracycline resistance under the control of the tryptophan promoter-operator. Thus the ligation mixture was transformed into E. coli 294 and colonies selected on LB plates containing 5 μ g/ml tetracycline.

Plasmid pHGH 207 was EcoRI digested and the trp promoter containing EcoRI fragment recovered by PAGE followed by electroelution. Plasmid pBRHI was EcoRI digested and the cleaved ends treated with bacterial alkaline phosphatase ("BAP") (1 μ g, in 50 mM tris pH 8 and 10 mM MgCl₂ for 30 min. at 65°C) to remove the phosphate groups on the protruding EcoRI ends. Excess bacterial alkaline phosphatase was removed by phenol extraction, chloroform extraction and ethanol precipitation. The resulting linear DNA, because it lacks phosphates on the protruding ends thereof, will in ligation accept only inserts whose complementary stick ends are phosphorylated but will not itself recircularize, permitting more facile screening for plasmids containing the inserts.

The EcoRI fragment derived from pHG 207 and the linear DNA obtained from pBRH1 were combined in the presence of T_4 ligase as previously described and ligated. A portion of the resulting mixture was transformed into *E. coli* strain 294 as previously described, plated on LB media containing 5 μ g/ml of tetracycline, and 12 tetracycline resistant colonies selected. Plasmid was isolated from each colony and examined for the presence of a DNA insert by restriction endonuclease analysis employing EcoRI and XbaI. One plasmid containing the insert was designated pHKY1.

The plasmid pHKY10, described above, is a derivative of pBR322 which contains a Bgl II site between the tetracycline resistance (Tc^R) promoter and structural gene (32). The large DNA fragment isolated after digesting pHKY10 with Pst I and Bgl II therefore contains part of the ampicillin resistance (Ap^R) gene and all of the Tc^R structural gene, but lacks the Tc^R promoter (Fig 6). The plasmid pGH6 (21) was digested with Eco RI, the resulting single stranded ends were filled in with DNA polymerase I, and the plasmid was cleaved with Pst I. The small fragment, containing part of the Ap^R gene, a double lac promoter and lac ribosome binding site, but lacking an ATG initiation triplet was isolated. A similar trp promoter fragment, containing the trp leader ribosome binding site, but lacking an ATG sequence (30), may be isolated from pHKY1 described above; see (32) (see Figure 6).

The trp fragment just referred to is an analog of the *E. coli* tryptophan operon from which the so-called trp attenuator has been deleted, See J. Bact. 133, 1457 (1978), to controllably heighten expression levels. Expression plasmids containing the modified trp regulon can be grown to predetermined levels in nutrient media containing additive tryptophan in quantities sufficient to repress the promoter-operator system, then be deprived of tryptophan so as to derepress the system and occasion the expression of the intended product.

The expression plasmids may be assembled via three part ligation reactions as shown in Figure 6. 15 ng (~ 0.05 pmole) of the assembled FIF gene (504 or 505 bp), 0.5 μ g (~ 0.2 pmole) of the large Pst I - Bgl II fragment of pHKY10 and 0.2 μ g (~ 0.3 pmole) of the appropriate promoter fragment were ligated and the mixture used to transform *E. coli* 294 (22). Plasmid DNA was prepared from individual transformants and analyzed by restriction mapping. Correct joining of the assembled gene to the promoter fragment should restore the Eco RI (lac) or the Xba I (trp) recognition sequences. The majority of the plasmids gave the expected

restriction enzyme digestion patterns. Individual clones (12 containing the trp promoter and 12 containing the lac promoter) were grown and extracts prepared for interferon assay as described in Materials and Methods.

When assayed on human amnion (WISH) cells for antiviral activity by the CPE inhibition assay (1) five of the trp transformants were positive (each approximately equivalent); eleven of the lac transformants gave equivalent IF activities. Therefore, one transformant from each series (pFIFlac9 and pFIFtrp69) was selected for further study (Table 1). DNA sequence analysis demonstrated that the desired attachment of promoter to FIF structural gene had occurred in both cases.

Table 1. Interferon activity in extracts of E. coli

| <u>E. coli</u> K-12 strain 294 transformed by: | Cell density (cells/ml) | IF Activity (units/l culture) | FIF molecules per cell |
|--|-------------------------|-------------------------------|------------------------|
| pBR322 | 3.5×10^8 | - | - |
| pFIF <u>lac</u> 9 | 3.5×10^8 | 9.0×10^6 | 2,250 |
| pFIF <u>trp</u> 69 | 3.5×10^8 | 1.8×10^7 | 4,500 |
| pFIF <u>trp</u> ³ 69 | 3.5×10^8 | 8.1×10^7 | 20,200 |

Cells were grown and extracts prepared as described in Materials and Methods. The human amnion (WISH) cell line was used for the CPE inhibition assay (1). Activities given are the average for three independent experiments. To determine the number of IF molecules per cell a FIF specific activity of 4×10^8 units/mg was used (2).

The amounts of fibroblast interferon produced by pFIFlac9 and pFIFtrp69 are shown in Table 1. The trp promoter gave a FIF expression level measurably higher than did the lac promoter. In an attempt to further increase FIF expression levels, pFIFtrp69 was cleaved with Eco RI and two 300 base pair Eco RI fragments containing the trp promoter (30) were inserted. The resulting plasmid, pFIFtrp³69, contains three successive trp promoters which read toward the FIF gene. The amount of FIF synthesized by E. coli K-12 strain 294/pFIF trp³69 is 4-5 times that produced by pFIF trp 69 (Table 1). This is apparently due to the derepression of the trp promoter which occurs when trp repressor levels are titrated by the multiple copies of the trp operator.

The FIF produced by E. coli K-12 strain 294/pFIFtrp69 behaves like

authentic human FIF. As shown in Table 2, its antiviral activity is about 30 times greater on human cells than bovine cells. In addition, the bacterially produced FIF is stable to treatment at pH 2 overnight and is not neutralized by rabbit antihuman leukocyte interferon antibodies (Table 3).

Table 2. Interferon activities measured on different cell types

| Cells | Interferon Activity (units/ml) | | |
|---------------|--------------------------------|--------|---|
| | LeIF | FIF | E. coli K-12 strain 2947pFIFtrp69 extract |
| Human amnion | 20,000 | 10,000 | 1280 |
| Bovine kidney | 13,000 | 400 | 40 |

LeIF and FIF were NIH standard solutions having 20,000 units/ml and 10,000 units/ml respectively. Assays were performed as described in Materials and Methods.

Table 3. Comparison of activities of extracts from E. coli K-12 strain 294/pFIFtrp69 with standard human leukocyte and fibroblast interferons

| | Interferon Activity (units/ml) | | |
|----------------------------------|--------------------------------|------|-----------------------------------|
| | LeIF | FIF | E. coli K-12 strain 2947pFIFtrp69 |
| untreated | 1000 | 1000 | 1000 |
| pH2 | 1000 | 1000 | 1000 |
| rabbit antihuman LeIF antibodies | <16 | 1000 | 1000 |

Experimental procedures described in Materials and methods. Assayed by CPE inhibition using WISH cells/Sindbis virus.

Purification

The purification procedure for bacterial derived fibroblast is as follows:

1. Frozen cells are suspended in twelve times volume per weight with sucrose lysis buffer (100mM Tris-HCl, 10 percent sucrose, 0.2M NaCl, 5mM EDTA, 0.2mM PMSF, pH 7.9) containing lysozyme at 1mg ml⁻¹. The cell suspension is stirred for 1 hour at 4°C and centrifuged. Fibroblast interferon activity remains in the supernatant.

2. Polyethyleneimine (5 percent v/v) is added to the sonicated supernatant to a final concentration of 0.5 percent (v/v). The solution is stirred for 1 hour at 4°C and centrifuged. Interferon activity remains in the supernatant.
3. Solid ammonium sulfate is added to the polyethyleneimine supernatant to a final concentration of 50 percent saturation, stirred for 30 minutes at 4°C and centrifuged. Interferon activity is in the 50 percent pellet.
4. The 50 percent ammonium sulfate pellet is suspended in one half the volume of the 50 percent ammonium sulfate suspension with Phosphate Buffered Saline (20 mM sodium phosphate 0.15M NaCl, pH 7.4). Polyethylene glycol 6000 (50 percent w/v in PBS) is added to a final concentration of 12 1/2 percent (v/v), stirred at 4°C for 2 hours and centrifuged. Interferon activity is in the pellet. The pellet is suspended in a minimal volume of sucrose lysis buffer and clarified by centrifugation.

This initial extraction procedure results in a purification of fibroblast interferon from 0.001 percent of the total protein to 0.05 percent of the total protein. This material can be further purified to homogeneity by the following column chromatography steps:

5. Affinity chromatography on Amicon Blue B in sucrose lysis buffer.
6. Anion exchange chromatography on QAE Sephadex in sucrose lysis buffer in the absence of 0.2M NaCl.
7. Size exclusion chromatography on Sephadex G-75 in sucrose lysis buffer.
8. Reverse phase high pressure liquid chromatography.

Parenteral Administration

FIF may be parenterally administered to subjects requiring antitumor or antiviral treatment. Dosage and dose rate may parallel that currently in use in clinical investigations of human derived materials, e.g., about $(1-10) \times 10^6$ units daily, and in the case of materials of purity greater than 1 percentage, likely up to, e.g., 150×10^6 units daily. Dosages of bacterially obtained FIF could be significantly elevated for

greater effect owing to the essential absence of human proteins other than FIF, which proteins in fibroblast-derived materials may act as pyrogens, exhibiting adverse effects, e.g., malaise, temperature elevation, etc.

As one example of an appropriate dosage form for essentially homogeneous bacterial FIF in parenteral form, 3 mg. FIF of specific activity of, say, 2×10^8 U/mg may be dissolved in 25 ml. 5 percentage serum albumin (human) - USP, the solution passed through a bacteriological filter and the filtered solution aseptically subdivided into 100 vials, each containing 6×10^6 units pure interferon suitable for parenteral administration. The vials are preferably stored in the cold (-20°C) prior to use.

The compounds of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby the polypeptide hereof is combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation are described in Remington's Pharmaceutical Sciences by E.W. Martin, which is hereby incorporated by reference. Such compositions will contain an effective amount of the interferon protein hereof together with a suitable amount of vehicle in order to prepare pharmaceutically acceptable compositions suitable for effective administration to the host. One preferred mode of administration is parenteral.

BIBLIOGRAPHY

- a 1. Stewart, W.E. II (1979) The Interferon System, Springer, New York.
- a 2. Knight, E. Jr. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 520-523.
- a 3. Berthold, W., Tan, C., and Tan, Y.H. (1978) J. Biol. Chem. 253, 5206-5212.
- a 4. Knight, E. Jr., Hunkapillar, M.W., Korant, B.D., Hardy, R.W.F. and Hood, L.E. (1980) Science 207, 525-526.
- a 5. Houghton, M., Stewart, A.G., Doel, S.M., Emtage, J.S., Eaton, M.A.W., Smith, J.C., Patel, T.P., Lewis, H.M., Porter, A.G., Birch, J.R., Cartwright, T., and Carey, N.H. (1980) Nucleic Acids Res. 8, 1913-1931.
- a 6. Taniguchi, T., Sakai, M., Fujii-Kuriyama, Y., Muramatsu, M., Kobayashi, S., and Sudo, T. (1979) Proc. Jan Acad. 855, 464-469.
- a 7. Derynck, R., Content, J., DeClercq, E., Volckaert, G., Tavernier, J., Devos, R., and Fiers, W. (1980) Nature 285, 542-547.
- a 8. Clewell, D.B. (1972) J. Bacteriol. 110, 667-676.
- a 9. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
- a 10. Taylor, J.M., Illemensee, R. and Summer, S. (1976) Biochim. Biophys. Acta 442, 324-330.
- a 11. Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 3961-3965.
- a 12. Crea, R., Kraszewski, A., Hirose, T. and Itakura, K. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 5765-5769.
- a 13. Hirose, T., Crea, R. and Itakura, K. (1978) Tetrahedron Letters 28, 2449-2452.
- a 14. Crea, R. and Horn, T. (1980) Nucleic Acids Res. 8, 2331-2348.
- a 15. Pestka, S., McInnes, J., Havell, E.A. and Vilcek, J. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 3898-3901.
- a 16. Green, M., Zehavi-Willner, T., Graves, P.N., McInnes, J. and Pestka, S. (1975) Arch. Biochem. Biophys. 172, 74-89.
- a 17. Gurdon, J.B., Lane, C.D., Woodland, H.R. and Marbaix, G. (1971) Nature 233, 177-182.
- a 18. Cavalieri, R.L., Havell, E.A., Vilcek, J. and Pestka, S. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 3287-3291.
- a 19. Scheller, R.H., Dickerson, R.E., Boyer, H.W., Riggs, A.D. and Itakura, K. (1977) Science 196, 177-180.
- a 20. Chang, A.C.Y., Nunberg, J.H., Kaufman, R.J., Erlich, H.A., Schinke, R.T. and Cohen, S.N. (1978) Nature 275, 617-624.
- a 21. Goeddel, D.V., Heyneker, H.L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D.G., Ross, M.J., Miozzari, G., Crea, R. and Seeburg, P.H. (1979) Nature 281, 544-548.

- Q 422. Backman, K., Ptashne, M. and Gilbert, W., (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 4174-4178.
- Q 423. Hershfild, V., Boyer, H.W., Yanofsky, C., Lovett, M.A. and Helinski, D.R. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 3455-3459.
- Q 424. Galau, G.A., Britten, R.J. and Davidson, E.H. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 1020-1023.
- Q 425. Birnboim, H.C. and Doly, J. (1979) Nucleic Acid Res. 7, 1513-1523.
- Q 426. Kafatos, F.C., Jones, C.W. and Efstratiadis, A. (1979) Nucleic Acids Res. 7, 1541-1552.
- Q 427. Denhardt, D.T. (1966) Biochem. Biophys. Res. Comm. 23, 641.
- Q 428. Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A.D. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 106-110.
- Q 429. Miller, J.H. (1972) Experiments in Molecular Genetics, pp.431-433, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Q 430. U.S. patent application of David V. Goeddel and Sidney Pestka U.S.S.N. 184909, filed September 8, 1980 for Microbial Production of Mature Human Leukocyte Interferon, Attorney docket 1980/104, assigns Genentech, Inc. and Hoffmann-LaRoche, Inc.
- Q 431. Taniguchi, T., Ohno, S., Fujii-Kuriyama, Y. and Muramatsu, M. (1980) Gene 10, 11-15.
- Q 432. U.S. patent application Serial No. 06/133,296 filed March 24, 1980 by Dennis G. Kleid et al., assigned to Genentech, Inc.
- Q 433. Klenow, H. and Henningsen, I. (1970) Proc. Natl. Acad. Sci. U.S.A. 65, 168-171.

Doeddel 27 ad.

374311

6/30/89

CLAIMS

1. A polypeptide comprising the amino acid sequence of a mature human fibroblast interferon, microbially produced and unaccompanied by any corresponding presequence or portion thereof.
2. A polypeptide according to Claim 1, unaccompanied by associated glycosylation.
3. The polypeptide according to Claim 1, optionally containing the amino acid methionine as the ordinarily first amino acid of said interferon.
4. The polypeptide according to Claim 1, optionally containing a cleavable conjugate or microbial signal protein attached to the N-terminus of the ordinarily first amino acid of said interferon.
5. A DNA sequence comprising a sequence coding for the polypeptide according to Claims 1, 3 or 4.
6. The DNA sequence according to Claim 5 operably linked with a DNA sequence capable of effecting microbial expression of a polypeptide according to Claims 1, 3 or 4.
7. A replicable microbial expression vehicle capable, in a transformant microorganism, of expressing a polypeptide according to Claims 1, 3 or 4.
8. A microorganism transformed with the vehicle according to Claim 7.
9. The microorganism according to Claim 8, obtained by transforming an E. coli strain.
10. A plasmid selected from the group consisting of pFIF lac 9, pFIF trp 69, and pFIF trp³ 69.
11. A transformed microorganism selected from the group consisting of an E. coli strain transformed independently with each of the plasmids according to Claim 10.
12. A transformed microorganism according to Claim 11 wherein said E. coli strain is E. coli K-12 strain 294.

530
351

536
27

435
320

435
252.33

435
320

435
252.33

13. A composition of matter comprising a therapeutically active fraction of a polypeptide consisting essentially of the amino acid sequence of a mature human fibroblast interferon, the balance of said composition comprising soluble microbial protein from which said polypeptide may be purified to a degree sufficient for effective therapeutic application.
14. A bacterial extract comprising greater than about 95 percent pure polypeptide consisting essentially of the amino acid sequence of a mature fibroblast interferon according to Claims 1, 2, 3, or 4.
15. A composition comprising a therapeutically effective amount of a mature human fibroblast interferon according to Claims 1, 3 or 4, suitable for pharmaceutical administration.
16. The composition according to Claim 15 suitable for parenteral administration.
17. A culture of microbial cells capable of producing a human fibroblast interferon in mature form.
18. The use of a mature human fibroblast interferon according to Claims 1, 3 or 4, for antitumor or antiviral treatment or for preparing pharmaceutical compositions useful for such treatment.
19. A process which comprises microbially expressing a human fibroblast interferon in mature form.
20. A process which comprises microbially expressing a polypeptide according to Claims 1, 3 or 4.
21. A process for producing a polypeptide according to Claims 1, 3 or 4 comprising causing a culture of a microorganism, transformed with a replicable microbial expression vehicle capable of expressing said polypeptide, to grow up and effect expression of said polypeptide and recovering said polypeptide.
22. A process for producing an expression vehicle according to Claim 7 comprising constructing a first DNA sequence coding for said polypeptide (defined in Claim 7) and operably linking said first DNA sequence with a second DNA sequence capable of effecting microbial expression of said first DNA sequence.

435
69.51

23. The process according to Claim 22 wherein said second DNA sequence comprises multiple tryptophan promoter-operators.

24. A product of the process according to Claims 19, 20, 21 or 22.

add B'

add c⁴

add E₁

ABSTRACT

A cDNA library is constructed using mRNA from human fibroblasts induced with poly(I):poly(C). A bacterial clone containing fibroblast interferon cDNA sequences identified by hybridization to a cDNA probe synthesized using deoxyoligonucleotide primers which hybridize to fibroblast interferon mRNA specifically. Expression plasmids are constructed which permit the synthesis in E. coli of 8×10^7 units of human fibroblast interferon per liter of culture. The bacterially produced fibroblast inteferon is indistinguishable from authentic human fibroblast interferon by several criteria.

0/374311

| Protein | 1 | 2 | 3 | 4 | |
|---------------------------|---|-----|-----|-----|-------------------|
| | Met | Ser | Tyr | Asn | |
| mRNA | (5') AUG-UC ^G _A -UA ^U _C -AA ^U _C | | | | (16 combinations) |
| | (5') AUG-AG ^U _C -UA ^U _C -AA ^C _U | | | | (8 combinations) |
| Complementary DNA primers | ATT- ^A _G TA- ^T _C GA-CAT | | | | Pool 1 |
| | ATT- ^A _G TA- ^A _G GA-CAT | | | | Pool 2 |
| | ATT- ^A _G TA- ^A _G CT-CAT | | | | Pool 3 |
| | GTT- ^A _G TA- ^T _C GA-CAT | | | | Pool 4 |
| | GTT- ^A _G TA- ^A _G GA-CAT | | | | Pool 5 |
| | GTT- ^A _G TA- ^A _G CT-CAT | | | | Pool 6 |

Figure 1.



Figure 2.

Print Of Drawing
As Original Filed

07/374311

3' =
S1 =
Met thr arg lys cys leu leu ala ile ala leu leu leu cys phe ser thr thr ala leu ser met ser tyr asn
ATG ACC AAC AAG TGT CTC CTC CAA ATT GCT CTC CTG TTG TGC TTC TCC ACT ACA GCT CTT TCC ATG AGC TAC AAC
10 20
LEU LEU GLY PHE LEU GLN ARG SER SER ASN PHE GLN CYS GLN LYS LEU LEU TRP GLN LEU ASN GLY ARG LEU GLU
TTG CTT GGA TTC CTA CAA AGA AGC AGC AAT TTT CAG TGT CAG AAG CTC CTG TGG CAA TTG AAT GGG AGG CTT GAA
130
30 40
TYR CYS LEU LYS ASP ARG MET ASN PHE ASP ILE PRO GLU GLU ILE LYS GLN LEU GLN GLN PHE GLN LYS GLU ASP
TAT TGC CTC AAG GAC AGS ATG AAC TTT GAC ATC CCT GAG GAG ATT AAG CAG CTG CAG CAG TTC CAG AAG GAG GAC
200
60 70
ALA ALA LEU THR ILE TYR GLU MET LEU GLN ASN ILE PHE ALA ILE PHE ARG GLN ASP SER SER SER THR GLY TRP
GCC GCA TIG ACC ATC TAT GAG ATG CTC CAG AAC ATC TTT GCT ATT TTC AGA CAA GAT TCA TCT AGC ACT GGC TGG
250 300
80 90 100
ASN GLU THR ILE VAL GLU ASN LEU LEU ALA ASN VAL TYR HIS GLN ILE ASN HIS LEU LYS THR VAL LEU GLU GLU
AAT GAG ACT ATT GTT GAG AAC CTC CTG GCT AAT GTC TAT CAT CAG ATA AAC CAT CTG AAG ACA GTC CTG GAA GAA
350
110 120
LYS LEU GLU LYS GLU ASP PHE THR ARG GLY LYS LEU MET SER SER LEU HIS LEU LYS ARG TYR TYR GLY ARG ILE
AAA CTG GAG AAA GAA SAT TTT ACC AGS GGA AAA CTC ATG AGC AGT CTG CAC CTG AAA AGA TAT TAT GGG AGG ATT
400 450
130 140 150
LEU HIS TYR LEU LYS ALA LYS GLU TYR SER HIS CYS ALA TRP THR ILE VAL ARG VAL GLU ILE LEU ARG ASN PHE
CTG CAT TAC CTG AAG GCC AAG GAG TAC AGT CAC TGT GCC TGG ACC ATA GTC AGA GTG GAA ATC CTA AGG AAC TTT
500
160 166
TYR PHE ILE ASN ARG LEU THR GLY TYR LEU ARG ASN END
TAC TTC ATT AAC AGA CTT ACA GGT TAC CTC CGA AAC TGA AGATCTCTAGCCTGTCCCTCTGGGACTGGACAATTGCTTCAGCA
550 600
TTCTCAACCAGCAGATGCTGTTAAGTGACTGATGGCTAATGTACTGCAATGAAGGACACTAGAAATTITGAAATTTTATTAAATTATGAGTT
650 700
ATTTTATTTTATTAAATTTTATTTTGGAAAATAAATTTTGGTGCAAAA
750
3'

Figure 3.

07/374311

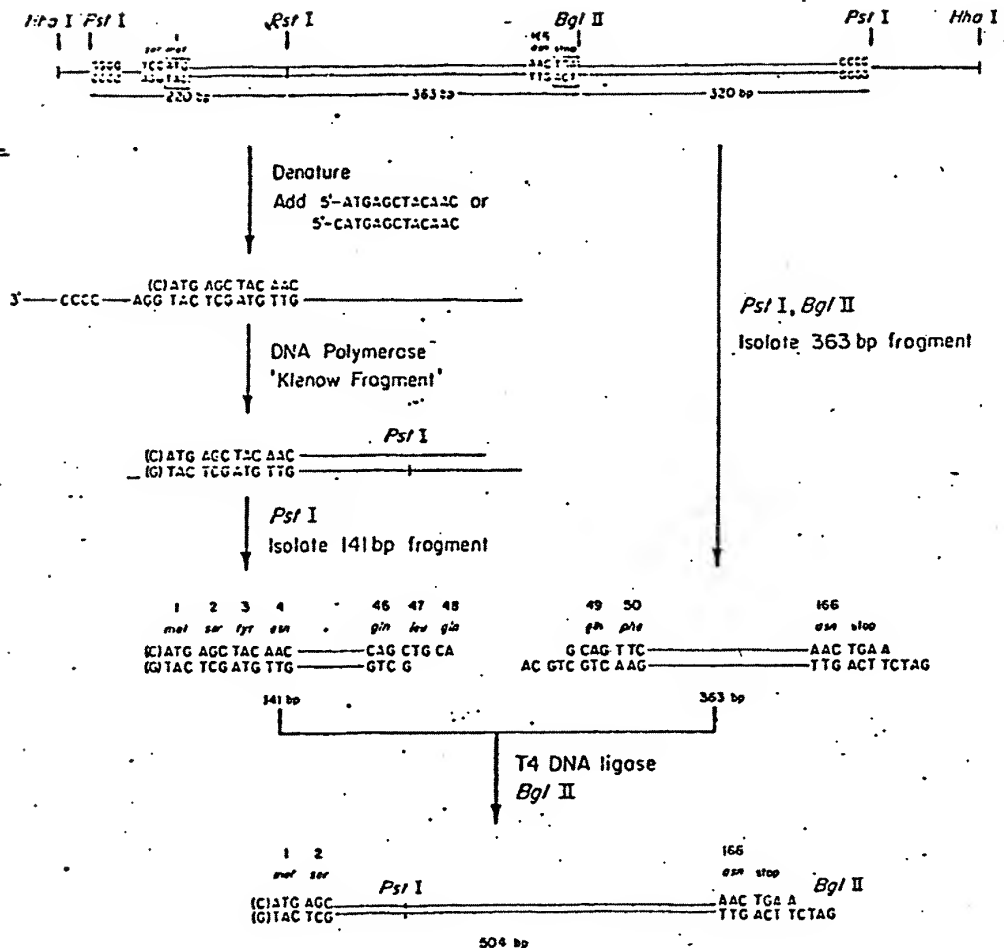


Figure 4.

07/274311

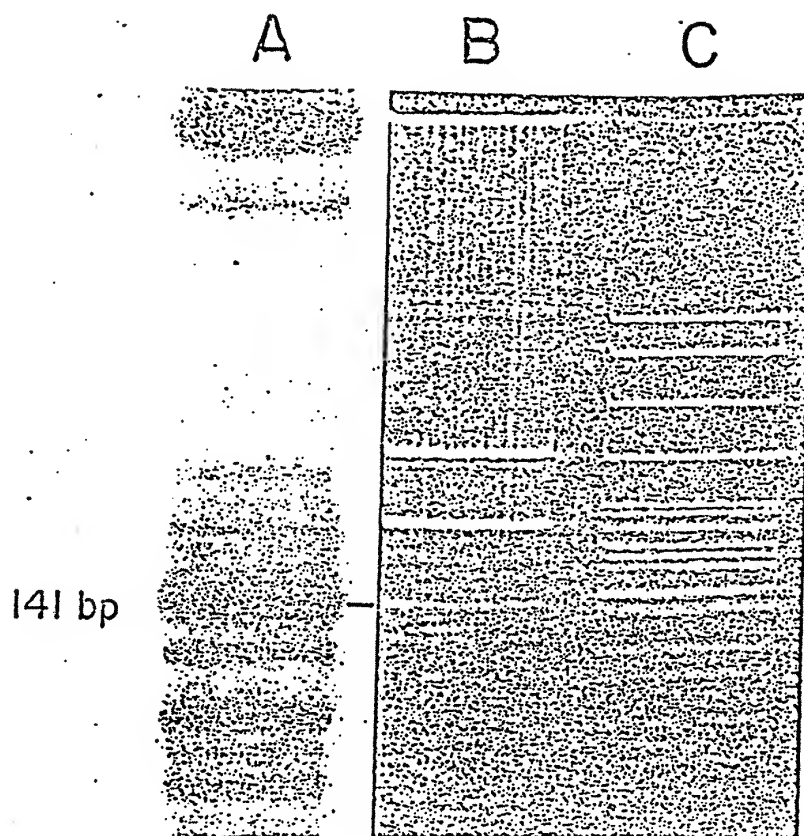


Figure 5.

07/374311

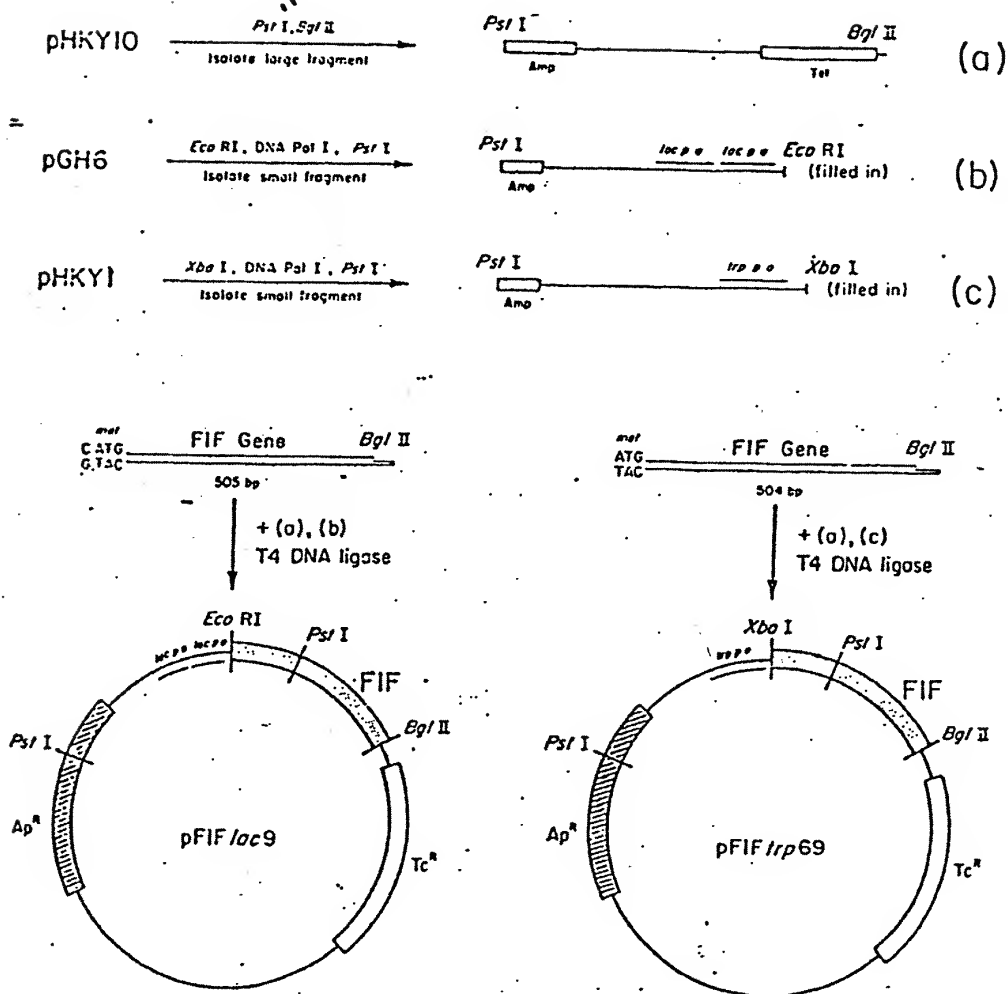


Figure 6.

TAB SS

PURIFICATION, AMINO ACID COMPOSITION, AND N-TERMINAL AMINO ACID SEQUENCE OF HUMAN FIBROBLAST INTERFERON

Ernest Knight, Jr.

Central Research and Development Department
E. I. du Pont de Nemours & Co.,
Wilmington, Delaware 19898

Michael Hunkapiller

Division of Biology
California Institute of Technology
Pasadena, California 91125

The chemical characterization of interferons is necessary in order to understand its actions at the molecular level. Characterization has been slow since only microgram quantities have been available of any interferon. Interferons from human fibroblast cells,^{1,2} human lymphoblastoid cells,³ human leukocytes,⁴ and mouse cells⁵⁻⁸ have recently been purified to homogeneity and some limited structural information has been reported.^{9,10} A major goal of the structural studies in interferons is the determination and comparison of amino acid sequences.

We report in this paper an improved procedure for the purification of human fibroblast interferon (HF-IF) that has been used to provide interferon for amino acid sequencing. Using an automated protein microsequencing technique previously described,^{11,12} we have determined the sequence of the 13 amino acid residues at the N-terminus of the HF-IF prepared by this method. We also report an amino acid composition.

Human diploid fibroblast cells (FS-4) were cultured and HF-IF was produced as previously described.¹ Interferon was assayed by a microassay technique,¹³ with vesicular stomatitis virus as the challenge virus. Interferon units are given in National Institutes of Health human fibroblast interferon units.

The crude HF-IF, 10-15 liters produced in the absence of serum, is made 1 M in NaCl and passed at room temperature through a column (4 × 10 cm) of Blue Sepharose (Pharmacia, Inc.) equilibrated with 0.02 sodium phosphate buffer, pH 7.2, containing 1 M NaCl. The HF-IF is retained while greater than 95% of the total protein passes through the column. The interferon is eluted with a 1:1 mixture of the column buffer and ethylene glycol, and each fraction is diluted immediately with 0.5 volume of the buffer (FIGURE 1a). Fractions containing interferon activity are pooled, diluted with two volumes of the column buffer, and passed through a small (1 × 6 cm) Blue Sepharose column for concentration. The interferon is eluted as described above (FIGURE 1b).

Fractions containing HF-IF are pooled, dialyzed against 1 mM Tris-HCl, pH 6.8, containing 0.02% sodium dodecyl sulfate (SDS, BioRad electrophoresis grade), and concentrated to dryness in a vacuum centrifuge. The HF-IF is then subjected to electrophoresis on a SDS-polyacrylamide slab gel and the HF-IF eluted as previously described¹ (FIGURE 2a). Fractions eluted from the gel are assayed for interferon activity (FIGURE 2a). Approximately 0.2 µg of HF-IF from the peak activity fraction is subjected to electrophoresis in this system again, and the gel is stained with Coomassie Blue (FIGURE 2b).

385

0077-8923/80/0350-0385 \$01.75/0 © 1980, NYAS

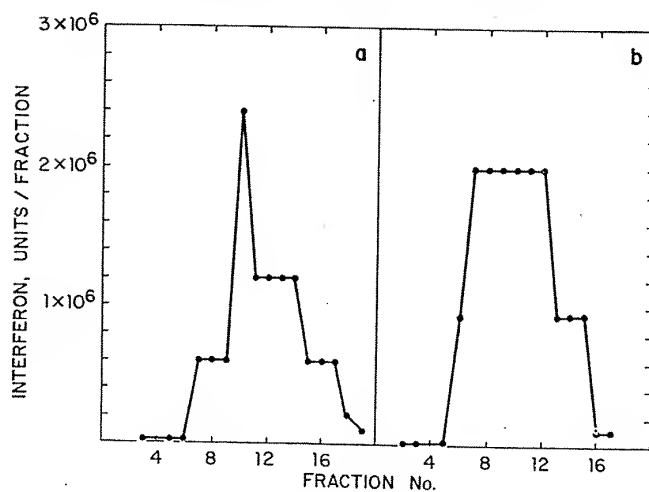


FIGURE 1. (a) Fractionation of crude interferon on a large column of Blue Sepharose. Elution of interferon with 50% ethylene glycol in column buffer begins at fraction 1. (b) Small Blue Sepharose column. Fractions 7-17 in (a) were pooled, passed through the small column, and eluted with 50% ethylene glycol in column buffer (fractions 1-20).

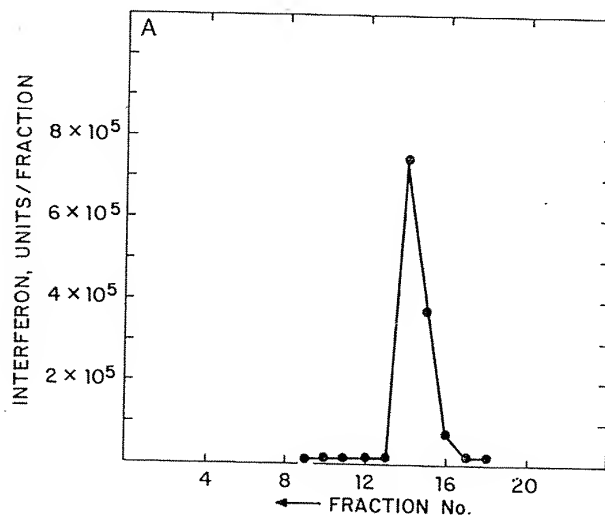


FIGURE 2a. Preparative electrophoresis of interferon, activity profile. Fractions 6-15 in FIGURE 1b were pooled, concentrated, and subjected to electrophoresis in a polyacrylamide slab gel, 0.75 mm thick. Fractions 14 and 15 (FIGURE 2a) were pooled and processed for amino acid sequencing.

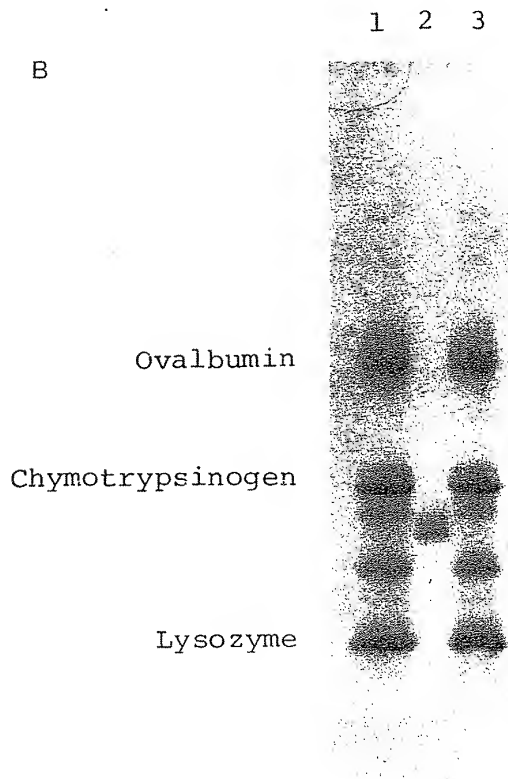


FIGURE 2b. Polyacrylamide slab gel, staining of proteins eluted from preparative gel in FIGURE 2a. Approximately 2% of the protein in fractions 14 and 15 (FIGURE 2a) was subjected to electrophoresis and stained.

The preparative electrophoresis fractions containing HF-IF are pooled and centrifuged for 30 min at 30,000 rpm, 4°C, to remove polyacrylamide gel particles. The interferon solution is dialyzed first against 0.15 M NaCl containing 0.1% SDS and then against 0.02% SDS, and it is concentrated to dryness in a vacuum centrifuge.

This purification procedure is simpler and shorter than that described previously.¹ Recoveries from the large Blue Sepharose column have ranged from 50–100%, and those from the small column approach 100%. The interferon (5×10^7 units/mg) eluted from these columns is stable for at least four weeks at 4°C in 1 M NaCl, 35% ethylene glycol, pH 7.2. Recoveries of activity from the SDS gels have ranged from 5–20%, and specific activities of this protein have ranged from $2\text{--}9 \times 10^8$ units/mg. Accurate specific activities are difficult to determine, and two- to fourfold differences above 1×10^8 units/mg are probably

TABLE 1
AMINO ACID COMPOSITION OF HUMAN FIBROBLAST INTERFERON

| | Mol % | Residues per 20,000 Daltons |
|------|-------|--------------------------------|
| Asp | 11.1 | 18.9 |
| Thr | 4.0 | 6.8 |
| Ser | 6.2 | 10.5 |
| Glu | 15.9 | 27.0 |
| Pro | 1.6 | 2.7 |
| Gly* | 4.6 | 7.8 |
| Ala | 5.9 | 10.0 |
| Cys† | 1.0 | 1.7 |
| Val | 3.5 | 6.0 |
| Met | 1.7 | 2.9 |
| Ile | 5.3 | 9.0 |
| Leu | 12.0 | 20.4 |
| Tyr | 4.4 | 7.5 |
| Phe | 5.5 | 9.4 |
| His | 2.9 | 4.9 |
| Lys | 6.8 | 11.6 |
| Arg | 6.4 | 10.9 |
| Trp‡ | 0.6 | 1.0 |

*Includes correction for free glycine present in unhydrolyzed protein.

†Determined after performic acid oxidation.

‡Determined after hydrolysis with mercaptoethane-sulfonic acid.

not meaningful. Overall yields of purified HF-IF from 10-15 liter batches of crude material ($5-7 \times 10^7$ total units, 8×10^4 units/mg) have averaged around 10 percent. This gives 5-20 μ g of homogeneous interferon.

Amino acid analysis on 1-2 μ g aliquots was performed on a Durrum D-500 amino acid analyzer (TABLE 1). Automated Edman degradation on 0.4-2 μ g aliquots of the purified interferon was performed in a spinning cup sequenator as previously described.^{11,12} Phenylthiohydantoin (Pth) amino acids were identified by high performance liquid chromatography on a Du Pont Zorbax CN column.¹⁴

The sequence of the N-terminal 13 amino acid residues of human fibroblast interferon determined by this microsequencing technique is given in FIGURE 3. Yields of Pth-Met at cycle 1 for three sequenator runs ranged from 60-100% (based upon protein quantitation by amino acid analysis), and the sequenator repetitive cycle yields were 92-95 percent. Any unblocked minor peptide sequence present at >5% of the reported sequence could have been detected by the methods used, but none has been observed. This result, coupled with the high initial Pth yields, confirms the homogeneity of the HF-IF polypeptide preparation.

| | | |
|---|---|----|
| 1 | 5 | 11 |
| Met - Ser - Tyr - Asn - Leu - Leu - Gly - Phe - Leu - Gln - Arg - Ser - Ser | | |

FIGURE 3. The N-terminal amino acid sequence of human fibroblast interferon.

This paper reports the amino acid composition and the N-terminal sequence of 13 amino acid residues of human fibroblast interferon.¹⁵ It is essential to determine the amino acid sequences of the interferons in order to identify an active site and to understand at the molecular level its mechanism of action. Moreover, determination of the amino acid sequences will allow comparison of structural features of interferons from different sources. This is an interesting possibility since interferons from different species and even from different cell types within an animal are thought to be different proteins.

REFERENCES

1. KNIGHT, E., Jr. 1976. *Proc. Nat. Acad. Sci. USA* 73: 520.
2. BERTHOLO, W., C. TAN & Y. H. TAN. 1978. *J. Biol. Chem.* 253: 5206.
3. ZOON, K. C., M. E. SMITH, P. J. BRIDGEN, D. ZUR NEUEN & C. B. ANFINSEN. 1980. *Proc. Nat. Acad. Sci. USA* 76: 5601.
4. RUBINSTEIN, M., S. RUBINSTEIN, P. C. FAMILLETTI, R. S. MILLER, A. A. WALMAN & S. PESTKA. 1979. *Proc. Nat. Acad. Sci. USA* 76: 640.
5. KNIGHT, E., Jr. 1975. *J. Biol. Chem.* 250: 4139.
6. YAMAMOTO, Y. & Y. KAWAOE. 1976. *J. Gen. Virol.* 33: 225.
7. KAWAKITA, M., B. CABRER, H. TAIRA, M. REBELLO, E. SLATTERY, H. WEIDELI & P. LENGUEL. 1978. *J. Biol. Chem.* 254: 598.
8. DEMAEYER-GUIGNARD, J., M. G. TOVEY, I. GRESSER & E. DEMAEYER. 1978. *Nature* 271: 622.
9. CABRER, B., H. TAIRA, R. J. BROEZE, T. D. KEMPLE, K. WILLIAMS, E. SLATTERY, W. H. KONIGSBERG & P. LENGUEL. 1979. *J. Biol. Chem.* 254: 3681.
10. TAN, Y. H., F. BARAKAT, W. BERTHOLD, H. SMITH-JOHANNSEN & C. TAN. 1979. *J. Biol. Chem.* 254: 8067.
11. HUNKAPILLER, M. W. & L. E. HOOD. 1978. *Biochemistry* 17: 2124.
12. BELL, J. R., M. W. HUNKAPILLER, L. E. HOOD & J. H. STRAUSS. 1978. *Proc. Nat. Acad. Sci. USA* 75: 2722.
13. ARMSTRONG, J. A. 1971. *Appl. Microbiol.* 21: 723.
14. JOHNSON, N. J., M. W. HUNKAPILLER & L. E. HOOD. 1979. *Anal. Biochem.* 100: 335.
15. KNIGHT, E., Jr., M. W. HUNKAPILLER, B. D. KORANT, R. W. F. HAROY & L. E. HOOD. 1980. *Science* 207: 525.

TAB TT

Construction of plasmids carrying the *cI* gene of bacteriophage λ

(repressor/DNA cloning/operon construction)

KEITH BACKMAN*, MARK PTASHNE†, AND WALTER GILBERT†

* Committee on Higher Degrees in Biophysics and † Department of Biochemistry and Molecular Biology, The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Contributed by Walter Gilbert, September 9, 1976

ABSTRACT By techniques of recombination *in vitro*, we have constructed a plasmid bearing the repressor gene (*cI*) of bacteriophage λ fused to the promoter of the *lac* operon. Strains carrying this plasmid overproduce λ repressor. This functional *cI* gene was reconstituted by joining DNA fragments bearing different parts of that gene. Flush end fusion techniques, involving no sequence overlap, were necessary for the construction; in certain cases, the abutting of the DNA molecules bearing ends generated by different restriction endonucleases creates a sequence at the junction which is recognized by one of the restriction endonucleases.

We have constructed a plasmid in which the promoter of the *lac* operon has been placed adjacent to the repressor gene (*cI*) of bacteriophage λ , by techniques of recombination *in vitro*. *Escherichia coli* strains carrying this plasmid overproduce repressor because transcription of *cI* originates mainly at the *lac* promoters and because each cell contains multiple copies of the plasmid. Certain novel aspects of our construction may be applied to the construction *in vitro* of other hybrid operons. Other λ repressor overproducing strains have been created by Groenborn and Müller-Hill (1), who have isolated *lac* promoter-*cI* fusions *in vivo*.

MATERIALS AND METHODS

E. coli K12 strain 294 (endo I^- , B_1^- , r_K^- , m_K^+ ; obtained from M. Meselson) was the host used in most of these experiments. A *lac* repressor overproducing strain (2), V2000 [*F'* *pro lac* i^{Q1Z^-} $U_{118}/\Delta(lac-pro)$ $Sm^R(\lambda imm^{21} plac5)$], was constructed and used in some experiments. The plasmid vectors pCR11 (3) and pMB9 (4) were used. Plasmid DNA was prepared by the method of Clewell and Helinski (5). DNA fragments generated by restriction endonucleases were prepared by electrophoresis on polyacrylamide gels as described by Maniatis *et al.* (6). *EcoRI*, *HindIII*, and *HaeIII* digestions were performed in *Hin* buffer (7) and *HpaII* digestions were performed in *Hpa* buffer (7). Reactions using *E. coli* DNA polymerase I (DNA nucleotidyltransferase, deoxynucleosidetriphosphate:DNA deoxynucleotidyltransferase, EC 2.7.7.7) (a gift of W. McClure) were performed in 50 mM Tris-HCl at pH 7.8, 5 mM $MgCl_2$, 1 mM 2-mercaptoethanol, 50 μ g/ml of bovine serum albumin plus 2 μ M each of dATP, dCTP, dGTP, and dTTP for 1 hr at 15° (8). T4 polynucleotide ligase [polynucleotide synthetase, poly(deoxyribonucleotide):poly(deoxyribonucleotide) ligase(AMP-forming), EC 6.5.1.1] reactions were performed in 6.6 mM Tris-HCl at pH 7.4, 6.6 mM $MgCl_2$, 6.6 mM 2-mercaptoethanol, 100 μ M ATP at 4°; DNA fragments were present at 100–300 μ g/ml. For joining DNA fragments bearing flush ends (9), 5-fold higher concentrations of ligase were used than were required to join an equivalent number of staggered ends, and the reactions were incubated at room temperature for 4 hr. Cells were transformed with DNA as described by Cohen *et al.* (10). Drug resistant transformants were selected on drug supplemented agar plates, and λ immune transformants were se-

lected on agar plates seeded with 10^9 λ KH54 (11) and 10^9 λ KH54h₈₀; in some cases, both selections were performed simultaneously. The immunity of plasmid containing strains was further tested by streaking single colonies across a streak of phage on an agar plate. Colonies of transformants which constitutively synthesized β -galactosidase appeared blue on agar plates containing a chromogenic, noninducing substrate, 5-chloro-4-bromo-3-indolyl- β -D-galactoside (40 μ g/ml). Assays of λ repressor (12), β -galactosidase (β -D-galactoside galactohydrolase, EC 3.2.1.23) (13), and the isolation of operator containing DNA fragments on nitrocellulose filters (14) have been described previously. Purified λ repressor was a gift of R. Sauer and purified *lac* repressor was a gift of A. Maxam. Experiments were carried out in a P1 (EK1) facility.

RESULTS

Preliminary considerations

We wished to proceed in two steps; first, to clone *cI* flanked by as little extraneous phage DNA as possible on a plasmid and then to insert a DNA fragment bearing the *lac* promoter near the beginning of *cI*. Two problems arose. No single restriction endonuclease cleaves λ DNA just outside the ends of *cI* without also cleaving within it, and no restriction endonuclease site suitable for inserting the *lac* promoter exists in λ DNA near the beginning of *cI*. We adopted a strategy based on the following considerations: gene *cI* can be neatly isolated on two DNA fragments, one of which bears two *HindIII* ends and the other of which bears one *HindIII* end and one *HaeIII* end (see Fig. 1). Proper joining of these fragments reconstitutes *cI* on a larger fragment bearing one *HindIII* end and one *HaeIII* end. *HindIII* ends are staggered and readily anneal to each other, whereas *HaeIII* ends are flush. Staggered ends which anneal to each other can be joined by T4 polynucleotide ligase (15), and flush ends can be joined to other flush ends by that enzyme (9). To clone our *cI* fragment, we sought a plasmid which could be opened so as to produce one *HindIII* end and one flush end. The plasmids we used, pCR11 and pMB9, each have a single *HindIII* site and a single *EcoRI* site. Although *EcoRI* produces staggered ends, we anticipated that these ends could be converted to flush ends by treatment with DNA polymerase I and the four deoxyribonucleotide triphosphates, because the recessed 3' end can be extended by copying the protruding 5' end of the complementary strand (a process we hereafter refer to as filling-in). Precise joining of a filled-in *EcoRI* end to a *HaeIII* end should produce a molecule recognized by *EcoRI* at the junction (see Fig. 2). This regenerated *EcoRI* site near the beginning of *cI* could then be used as a site to insert the *lac* promoter.

The plasmids pCR11 and pMB9 are derived from Col E1, which is normally present in about twenty copies per cell (16). pCR11 and pMB9 carry drug resistance determinants (kanamycin and tetracycline, respectively).

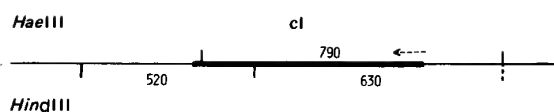


FIG. 1. Restriction endonuclease sites in the vicinity of the *cI* gene of λ . *Hae*III cleavage sites are indicated above, and *Hind*III sites below, the map of the region including and flanking *cI*. (The extreme right *Hae*III site is also indicated below the line.) Distances in base pairs between various cleavage sites are given. The extent of *cI* is indicated by the heavy line. The extreme right *Hae*III site is in the *cro* gene, and the extreme left *Hind*III site is in the gene which lies to the left of *cI*, *rex*. The arrow indicates the direction of transcription of *cI*.

Cloning *cI*

With these considerations in mind, we proceeded as follows (see Fig. 2): plasmid pCR11 DNA was digested with *Eco*RI and the protruding single-stranded ends converted to double-stranded flush ends by filling-in with DNA polymerase I. The product was then digested with *Hind*III to produce a vector DNA molecule bearing one *Hind*III end and one flush end. This procedure removes at least part of the kanamycin resistance gene(s). Two DNA fragments bearing portions of *cI* were produced as follows (see Fig. 1): λ DNA was digested with *Hae*III, and a 790 base pair fragment was isolated that carries the right portion of *cI* plus about 150 base pairs to the right of *cI*. This fragment was further digested with *Hind*III to produce a fragment 630 base pairs long. Separately, λ DNA was digested with *Hind*III and a 520 base pair fragment bearing the left portion of *cI* was isolated. Together these two fragments span the entire *cI* gene. Gene *cI* was then simultaneously reconstituted and inserted in the plasmid by treating a mixture of the specially prepared plasmid and the fragments which span *cI* with T4 polynucleotide ligase (see Fig. 2). Five λ immune clones were isolated from bacteria transformed with the products of this reaction, and their plasmids were analyzed by restriction endonuclease digestion. One isolate, pKB155, produced three fragments (5000, 630, and 520 base pairs long) when cleaved with *Eco*RI and *Hind*III, as predicted by the construction shown in Fig. 2 (E and F). In particular, the fused *Hae*III-*Eco*RI junction to the right of *cI* was cleaved again by *Eco*RI.

The *cI* gene was transferred from pKB155, which bears no drug resistance marker, to the plasmid pMB9, which carries a tetracycline resistance gene, as follows: an 1150 base pair piece of DNA bearing the entire *cI* gene was excised from pKB155 by complete digestion with *Eco*RI and partial digestion with

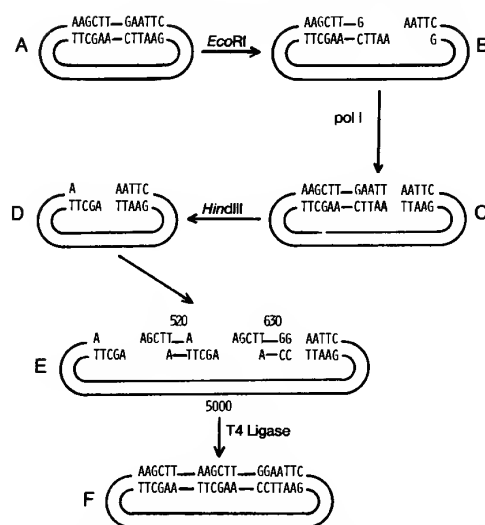


FIG. 2. Cloning the *cI* gene. Gene *cI*, carried on two DNA fragments, was inserted into the plasmid pCR11 as diagrammed (see text). The plasmid pCR11 (A) was digested with *Eco*RI (B), and the protruding 5' ends were filled in (see text) with DNA polymerase I (pol I) (C); the product was digested further with *Hind*III (D). The *Hae*III 790 base pair fragment digested with *Hind*III and the *Hind*III 520 base pair fragment (see Fig. 1) were added to the prepared plasmid (E). No attempt was made to remove the extra pieces of DNA generated by the *Hind*III digestions of the plasmid and of the *Hae*III 790 fragment. The mixture was treated with polynucleotide ligase, and the desired recombinant (F) was selected as described in the text.

*Hind*III. Plasmid pMB9 was digested with *Eco*RI and *Hind*III, which removes a nonessential 350 base pair fragment; the 1150 base pair *cI* gene fragment was added, and the mixture was treated with ligase. Bacteria were transformed with the products of ligation, and tetracycline resistant lambda immune clones were readily isolated. Restriction endonuclease analysis of several isolates (not shown), typical of which is pKB158, confirmed the structure shown in Fig. 3.

Construction of a plasmid fusing the *lac* promoter to the *cI* gene

A plasmid bearing a (nominally) 205 base pair fragment of DNA which carries the *lac* UV5 promoter-operator region was constructed by F. Fuller. He used a 203 base pair fragment of known sequence generated by *Hae*III. The *Hae*III ends of this

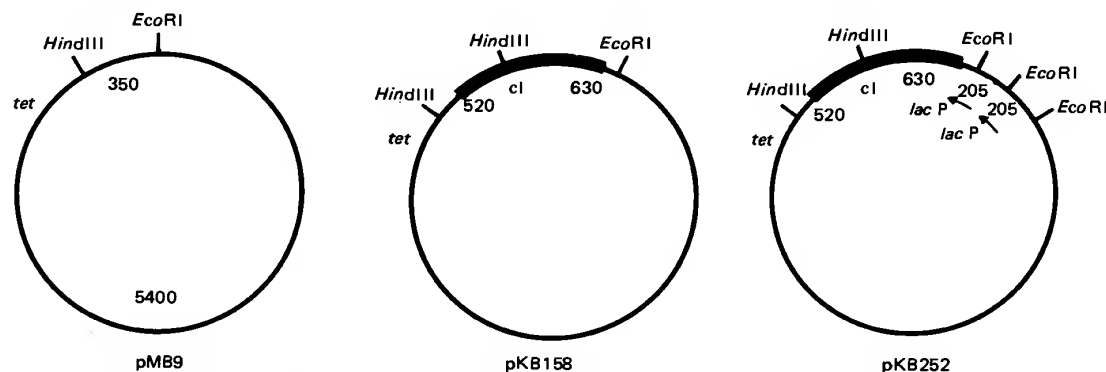


FIG. 3. The structure of pMB9, pKB158, and pKB252. The *Eco*RI and *Hind*III sites, the sizes of the cleavage products, and the location of *cI* in the various plasmids are shown. The location and orientation of the *lac* promoters in pKB252 are indicated.

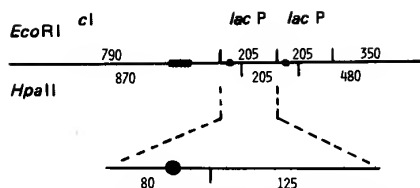


FIG. 4. Structure of pKB252 in the *lac* promoter-*cI* gene region. The location of *Eco*RI and *Hpa*II cleavage sites are shown, respectively, above and below the map of the 1550 base pair fragment produced by *Hae*III digestion of pKB252. The locations of the *lac* and lambda operators on this fragment are indicated by the solid circles and solid box, respectively. Distances between various endonuclease sites and the position of the *Hpa*II site in the *lac* promoter are given in base pairs.

fragment were converted to *Eco*RI ends by joining the fragment to filled-in *Eco*RI ends on a plasmid by using the methods described above. Colonies harboring plasmids which carry the *lac* promoter-operator were identified by their constitutive synthesis of β -galactosidase, which renders them blue on agar plates containing 5-chloro-4-bromo-3-indolyl- β -D-galactoside. The *lac* promoter fragment serves as a "portable promoter" with an *Eco*RI site 65 base pairs from the startpoint of transcription (corresponding to the ninth amino-acid residue of the *lac* α gene). This fragment was excised from its carrier with *Eco*RI and inserted near the beginning of *cI* at the single *Eco*RI site in pKB158. About one half of the tetracycline resistant clones which were isolated carried the *lac* promoter fragment on the plasmid, as manifested by their constitutive synthesis of β -galactosidase. All of these clones were immune to λ vir (unlike the strain carrying pKB158) and sensitive to λ mm⁴³⁴*cI*⁻ in cross-streak tests, although many clones exhibited some cell death at the junction of the cross-streak with λ vir. Three clones which exhibited no apparent sensitivity to λ vir were selected for further study. Typical of the plasmids in these clones is pKB252.

Structure of pKB252

An *Eco*RI plus *Hind*III digest of pKB252 yielded the 205 base pair promoter fragment as well as the fragments produced by a similar digestion of pKB158. When pKB252 was digested with *Hae*III, a 1550 base pair fragment was produced. This fragment, which extends from the *Hae*III site in *cI* to a *Hae*III site which is 350 base pairs beyond the *Eco*RI site in the plasmid DNA (17), was expected to contain the *lac* promoter. Analysis of this 1550 base pair fragment confirmed the structure of pKB252 as shown in Fig. 3. The products of *Hpa*II digestion of this fragment were 205, 480, and 870 base pairs long (see the diagram of Fig. 4). Because the *cI* region of λ does not have any *Hpa*II sites (11), and the *lac* promoter fragment does have a *Hpa*II site about 85 bases from one end, this pattern of fragments suggested the structure shown in Fig. 4. In particular, the 205 base pair product suggested the presence of multiple *lac* promoters. If only one promoter were present, then no such piece would have been produced. Moreover, all such promoters must have the same orientation; if any two promoter regions were oppositely oriented, then *Hpa*II digestion would have produced a 160 or a 250 base pair fragment.

The exact number of *lac* promoters in pKB252 was determined as follows: uniformly ³²P labeled plasmid DNA was digested with *Hae*III and the 1550 base pair fragment was isolated. This fragment was digested with *Eco*RI and the products separated on a polyacrylamide gel. The resulting fragments (205, 360, and 790 base pairs) were cut from the gel and the

Table 1. The number of *lac* promoter fragments in pKB252

| Fragment size in base pairs | Radioactivity ³² P counts per minute | Stoichiometry |
|--------------------------------|--|---------------|
| 790 | 3110 | 1 |
| 360 | 1345 | 1 |
| 205 | 1350 | 2 |

Stoichiometry is calculated by comparing the ratios of radioactivity/fragment size.

radioactivity in each fragment was determined (Table 1). The 205 base pair fragment was present in twice molar quantities compared to the 360 and 790 base pair fragments which indicated that there are two *lac* promoters in pKB252.

To prove that the *lac* promoters transcribe toward *cI*, we took advantage of the fact that *Hpa*II cuts near the *lac* operator in the *lac* promoter. The *Hae*III 1550 fragment was digested with *Hpa*II and the fragments produced were mixed with either *lac* or lambda repressor and passed through nitrocellulose filters. The fragments retained on the filters were eluted and analyzed on a polyacrylamide gel. It was found that the 870 base pair fragment was retained by both *lac* and λ repressors, the 205 base pair fragment was retained only by *lac* repressor, and the 480 base pair fragment was not retained by either repressor. These results are consistent only with the orientation shown in Fig. 4 in which the *lac* promoters transcribe toward *cI*.

Control of repressor synthesis *in vivo*

Cell extracts of strains containing various plasmids were assayed for λ repressor and, in some cases, an aliquot of the culture was assayed for β -galactosidase. The results are listed in Table 2. Strains carrying plasmids bearing *cI* but not the *lac* promoter (pKB155 and pKB158) contained about five times more repressor than did single lysogens. Strains carrying pKB252, which has two *lac* promoters, produced 35 times more repressor than did single lysogens.

Synthesis of lambda repressor in pKB252 strains is regulated by *lac* repressor (Table 2). A strain bearing a wild-type *lac* operon does not make enough *lac* repressor to repress significantly the synthesis of λ repressor from the *lac* promoters on pKB252 (Table 2). When this plasmid was transferred to a strain which

Table 2. Repressor synthesis in plasmid containing strains

| Strain | Repressor level $\times 10^{-3}$ | β -Galactosidase level |
|----------------------|-------------------------------------|---------------------------------|
| 28(λ) | 0.3 | N.D. |
| 294/pKB155 | 1.0 | N.D. |
| 294/pKB158 | 1.5 | 3 |
| 294/pKB252 | 9.7 | 2.8×10^3 |
| 294/pKB252 (+IPTG) | 10.7 | 2.8×10^3 |
| V2000/pKB252 | 1.3 | 10 |
| V2000/pKB252 (+IPTG) | 2.9 | 1.9×10^3 |

Various strains were grown overnight at 37° in M9 salts and glucose medium, and were assayed for lambda repressor and β -galactosidase. The results are given in units of repressor per milligram of protein. One unit is the amount of repressor which gives half-maximal binding in the nitrocellulose filter binding assay and corresponds to approximately 0.5 ng of repressor. β -Galactosidase results are given in the units of Miller (13). A single lysogen, 28 (λ), is included for comparison purposes. Strain 294 makes wild-type *lac* repressor levels, whereas strain V2000 over-produces *lac* repressor about 100-fold (2). IPTG, isopropyl-thiogalactoside. N.D., not determined.

Table 3. Use of the filling-in method with various restriction endonucleases

| Restriction endonuclease | Recognition sequence | Requirement for site regeneration | Subsite | Subsite Recognition sequence |
|--------------------------|----------------------------|-----------------------------------|----------------|------------------------------|
| <i>EcoRI</i> | ↓ GAATTC CTTAAG ↑ | 5' C | <i>EcoRI</i> * | ↓ AATT TTAA ↑ |
| <i>EcoRII</i> | ↓ CCAGG GGTCC ↑ | None | <i>EcoRII</i> | ↓ CCAGG GGTCC ↑ |
| <i>HindIII</i> | ↓ AAGCTT TTCGAA ↑ | 5' T | <i>AluI</i> | ↓ AGCT TCGA ↑ |
| <i>BamI</i> | ↓ GGATCC CCTAGG ↑ | 5' C | <i>MboI</i> | ↓ GATC CTAG ↑ |
| <i>BglIII</i> | ↓ AGATCT TCTAGA ↑ | 5' T | <i>MboI</i> | ↓ GATC CTAG ↑ |

Several endonucleases which produce 5' protruding ends are listed with the DNA sequence which they recognize (20–24, R. J. Roberts, G. Wilson, and F. E. Young, submitted for publication). After filling in these ends with DNA polymerase I, the original recognition sequence can be regenerated by joining the filled-in end to a flush ended DNA molecule which meets the requirement listed in the third column. When joined to a flush ended fragment not meeting that requirement, the filled-in end generates a sub-specificity, which is described in the fourth and fifth columns (20, 25, R. E. Gelinas, P. A. Myers, K. Murray, and R. J. Roberts, unpublished). The top line of each recognition sequence reads from the 5' to the 3' end; the bottom line reads in the opposite direction.

overproduces *lac* repressor, the synthesis of λ repressor was reduced to the basal (i.e., pKB158) level. IPTG (isopropyl-thiogalactoside), an inducer of the *lac* operon, partially induces β -galactosidase synthesis in cells which overproduce *lac* repressor. Table 2 shows that isopropyl-thiogalactoside also partially induces λ repressor synthesis when such cells carry pKB252.

DISCUSSION

We have described the use of techniques of DNA recombination *in vitro* to construct a plasmid on which the λ repressor gene (*cI*) is transcribed largely from the promoter of the *lac* operon. This plasmid has two copies of the *lac* promoter which transcribe toward *cI*. As predicted by this arrangement, the synthesis of λ repressor in strains bearing this plasmid is regulated by the *lac* repressor. The *lac* promoter fragments in pKB252 contain the startpoint of translation of the *lac z* gene; however, several translational stop signals immediately precede *cI* (18), and repressor synthesized in strains carrying pKB252 is not a fusion product initiated at the *lac z* translational startpoint. Strains carrying this plasmid produce about 35 times more repressor than do single lysogens, as measured by DNA binding activity; however, these levels are variable and are often as much as 3-fold higher. The source of this variation is not understood. These repressor levels are 10- to 25-fold lower than would be expected on the basis of the known strength of the *lac* promoter (19). This inefficient expression might occur because transcription beginning at the *lac* promoter attenuates fre-

quently before transcribing *cI*, because the progress of RNA polymerase is blocked by λ repressor bound to the operator just to the right of *cI*, or because the mRNA produced is inefficiently translated. The level of repressor was not higher when plasmid pKB252 was carried by a strain bearing an *SuA* mutation. Nevertheless, the high levels of repressor made by strains carrying pKB252 have greatly facilitated structural studies of repressor.

Strains carrying plasmids in which *cI* is transcribed from its own promoter produce about five times more repressor than does a single lysogen, even though these plasmids should be present in about 20 copies per cell (16). Gene *cI* is known to regulate its own synthesis (18), and these strains are exemplary of such autogenous regulation.

The expression of tetracycline resistance in strains carrying pKB158 and pKB252 may be under the control of the promoter for *cI* (P_{RM}) and the *lac* promoter. H. W. Boyer (personal communication) has found evidence suggesting that the *HindIII* site in pMB9 is within the promoter of the tetracycline resistance gene. In our experiments, tetracycline resistant λ immune clones were readily isolated which were missing the pMB9 specific material between the *EcoRI* and *HindIII* sites; presumably the expression of tetracycline resistance in these cases results from an extension of the *cI* mRNA.

Our construction required the use of new techniques which are generally applicable. We have assembled a hybrid operon from separate pieces; some junctions involved sequence overlap, some did not. Furthermore, the DNA polymerase I filling-in technique allowed us to join precisely DNA fragments which could not be joined directly by treatment with ligase. These techniques greatly expand the spectrum of restriction endonucleases which are useful in cloning experiments. An especially important feature of these techniques is the ability to regenerate endonuclease recognition sites at junctions involving filled-in ends. We demonstrate this explicitly in one case (the joining of a filled-in *EcoRI* end to a *HaeIII* end), and available sequence data suggest that it will be a general feature of the method (see Table 3).

W. G. is American Cancer Society Professor of Molecular Biology. We wish to thank F. Fuller and L. Johnsrud, who constructed and characterized the *lac* promoter containing plasmid, and R. Brent, who prepared the *lac* promoter fragment. We also wish to thank R. T. Sauer for λ repressor and for performing some of the repressor assays, A. Jeffrey for restriction endonucleases, A. Maxam for *lac* repressor, and W. McClure for DNA polymerase I. This work was supported by grants from the National Science Foundation and the National Institutes of Health.

1. Gronenborn, B. & Müller-Hill, B. (1976) *Mol. Gen. Genet.*, in press.
2. Müller-Hill, B., Fanning, T., Geisler, N., Gho, D., Kania, J., Kathman, P., Meissner, H., Schlotmann, M., Schmitz, A., Treisch, I. & Beyreuther, K. (1975) in *Protein Ligand Interactions*, eds. Sund, H. & Blauer, G. (Walter de Gruyter, Berlin), pp. 211–224.
3. Covey, C., Richardson, D. & Carbon, J. (1976) *Mol. Gen. Genet.* 145, 155–158.
4. Rodriguez, R. L., Bolivar, F., Goodman, H. M., Boyer, H. W. & Betlach, M. (1976) "Molecular mechanisms in control of gene expression," eds. Nierlich, D. P., Rutter, W. S. & Fox, C. F., in *Proceedings of the ICS-UCLA Symposia on Molecular and Cellular Biology* (Academic Press, San Francisco, California), Vol. 5, in press.
5. Clewell, D. B. & Helinski, D. R. (1970) *Biochemistry* 9, 4428–4440.
6. Maniatis, T., Jeffrey, A. & van deSande, H. (1975) *Biochemistry* 14, 3787–3794.

7. Maniatis, T., Ptashne, M. & Maurer, R. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 857-868.
8. Maniatis, T., Jeffrey, A. & Kleid, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184-1188.
9. Sgaramella, V., van deSande, J. & Khorana, H. G. (1970) *Proc. Natl. Acad. Sci. USA* **67**, 1468-1475.
10. Cohen, S. N., Chang, A. C. Y. & Hsu, L. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2110-2114.
11. Blattner, F. R., Fiandt, M., Hass, K. K., Twose, P. A. & Szybalski, W. (1974) *Virology* **62**, 458-471.
12. Chadwick, P., Pirrotta, V., Steinberg, R., Hopkins, N. & Ptashne, M. (1970) *Cold Spring Harbor Symp. Quant. Biol.* **35**, 283-294.
13. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.)
14. Maniatis, T. & Ptashne, M. (1973) *Nature* **246**, 133-136.
15. Sgaramella, V. & Khorana, H. G. (1972) *J. Mol. Biol.* **72**, 427-444.
16. Clewell, D. B. & Helinski, D. R. (1972) *J. Bacteriol.* **110**, 1135-1146.
17. Maniatis, T., Kee, S. G., Efstratiadis, A. & Kafatos, F. C. (1976) *Cell* **8**, 163-182.
18. Ptashne, M., Backman, K., Humayun, M. Z., Jeffrey, A., Maurer, R., Meyer, B. & Sauer, R. T. (1976) *Science*, **194**, 156-161.
19. Cohn, M. (1957) *Bacteriol. Rev.* **21**, 140-168.
20. Polisky, B., Greene, P., Garfin, D., McCarthy, B., Goodman, H. & Boyer, H. W. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3310-3314.
21. Bigger, C. H., Murray, K. & Murray, N. E. (1973) *Nature New Biol.* **244**, 7-10.
22. Boyer, H. W., Chow, L. T., Dugaiczky, A., Hedgpeth, J. & Goodman, H. M. (1973) *Nature New Biol.* **244**, 40-43.
23. Old, R., Murray, K. & Roizes, G. (1975) *J. Mol. Biol.* **92**, 331-339.
24. Pirrotta, V. (1976) *Nucleic Acids Res.* **3**, 1747-1760.
25. Roberts, R. J., Myers, P. A., Morrison, A. & Murray, K. (1976) *J. Mol. Biol.* **102**, 157-165.

TAB UU



US005460811A

United States Patent [19]**Goeddel et al.**[11] **Patent Number:** **5,460,811**[45] **Date of Patent:** **Oct. 24, 1995**[54] **MATURE HUMAN FIBROBLAST
INTERFERON**[75] Inventors: **David V. Goeddel; Roberto Crea,**
Burlingame, Calif.[73] Assignee: **Genentech, Inc.,** South San Francisco,
Calif.[21] Appl. No.: **365,284**[22] Filed: **Jun. 12, 1989****Related U.S. Application Data**[60] Continuation of Ser. No. 889,722, Jul. 28, 1986, abandoned,
which is a division of Ser. No. 291,892, Aug. 11, 1981,
abandoned, which is a continuation-in-part of Ser. No.
190,799, Sep. 25, 1980, abandoned.[51] Int. Cl.⁶ **C07K 15/26; A61K 37/66**[52] U.S. Cl. **424/85.6; 424/85.4; 530/351**[58] Field of Search **530/351; 424/85.6,**
424/85.4[56] **References Cited****U.S. PATENT DOCUMENTS**

| | | | |
|-----------|---------|---------------------|-----------|
| 4,262,090 | 4/1981 | Colby et al. | 435/97 |
| 4,289,689 | 9/1981 | Friesen et al. | 260/112 R |
| 4,332,892 | 6/1982 | Ptashne et al. | 435/172 |
| 4,342,832 | 8/1982 | Goeddel et al. | 435/172 |
| 4,418,149 | 4/1983 | Ptashne et al. | 435/172 |
| 4,874,702 | 10/1989 | Fiers et al. | 435/172 |
| 5,015,730 | 5/1991 | Friesen et al. | 530/351 |

FOREIGN PATENT DOCUMENTS

| | | |
|---------|--------|----------------------|
| 6694 | 1/1980 | European Pat. Off. . |
| 2007676 | 5/1979 | United Kingdom . |

OTHER PUBLICATIONSDarnell et al. (Darnell), *Molecular Cell Biology*, "Golgi
Vesicles: Sorting and Glycosylation of Secretory and Mem-
brane Proteins", Second Edition, pp. 661-667 (1990).Houghton et al., *Nature* 285:536 (published Jun. 19, 1980).Goeddel et al., *Nature* 287:411-416 (published Oct. 2,
1980).Horoszczewicz et al., *Cancer Treatment Rep.* 62:1899-1906
(1978)Stewart II et al., *Virology* 97:473-476 (1979).Pestka et al., *Proc. Nat'l Acad. Sci. USA* 72: 3898-3901
(1975).Reynolds et al., *Proc. Nat'l Acad. Sci. USA* 72: 4881-4885
(1975).Cavalieri et al., *Proc. Nat'l Acad. Sci. USA* 74: 4415-4419
(1977).Stewart II, in *The Interferon System* (Springer-Verlag, New
York, 1979), pp. 90-96, 335, 386 and 389.Grob et al., *Biochemistry* 18: 5782-86 (1979).Havell et al., *Virology* 63: 475-83 (1975).Houghton et al., *Nucleic Acids Res.* 8: 1913-31 (1980).Bose et al., *J. Biol. Chem.* 251: 1659-62 (1976).Davey et al., *Biochemistry* 15: 704-13 (1976).Eagon et al., *J. Biol. Chem.* 252: 2372-83 (1977).Mizrahi et al., *J. Biol. Chem.* 253: 7612-15 (1978).Stewart et al., *Proc. Natl. Acad. Sci. USA* 74: 4200-04
(1977).Struck et al., in *The Biochemistry of Glycoproteins and*
Proteoglycans, Plenum Press (W. J. Lennarz ed.), 1980, pp.
35-83.Taniguchi et al., *Proc. Japan Acad.* 55(Scr.B):464-469
(1979).Taniguchi et al., *Gene* 10:11-15 (published May 19, 1980).Taniguchi et al., *Proc. Nat'l Acad. Sci. USA* 77:4003-4006
(published Aug. 13, 1980).Taniguchi et al., *Proc. Nat'l Acad. Sci. USA* 77:5230-5233
(published Oct. 15, 1980).Taniguchi et al., *Nature* 285:547-549 (published Jun. 19,
1980).Derynck et al., *Nature* 285:542-547 (published Jun. 19,
1980).Derynck et al., *Nature* 287:193-197 (published Sep. 18,
1980).Guarente et al., *Science* 209:1428-1430 (published Sep. 19,
1980).Anonymous, *Research Disclosure* (Jul. 1979).Houghton et al., *Nucleic Acids Res.* 8(13):2885-2894 (pub-
lished Jul. 14, 1980).Goeddel et al., *Nucleic Acids Res.* 8(18):4057-4074 (pub-
lished Sep. 25, 1980).Goeddel et al., *Nature* 281:544-548 (1979)Chang et al., *Nature* 275:617-624 (1978).

Backman et al., Cell 13:65-71 (1978).
 Knight, Proc. Nat'l Acad. Sci. USA 73(21):520-523 (1976).
 Tan et al., J. Biol. Chem. 244:8067-8073 (1979).
 Berthold et al., J. Biol. Chem. 253(14):5206-5212 (1978).
 Knight et al., Science 207:525-527 (1980).
 Reynolds et al., Biochem. Biophys. Res. Comm. 63(1):107-112 (1977).
 Havell et al., J. Biol. Chem. 252(12):4425-4427 (1977).
 Cavalieri et al., Proc. Nat'l. Acad. Sci. USA 74:3287-3291 (1977).
 Pestka et al., Ann. Rev. New York Acad. Sci. 284:697-702 (1977).
 Content et al., Virology 122:466-470 (1982).
 Roberts et al., Proc. Nat'l. Acad. Sci. USA 76:760-764 (1979).
 Roberts et al., Proc. Nat'l. Acad. Sci. USA 76:5596-5600 (1979).
 Guarente et al., Cell, 20:543-553 (1980).
 Backman et al. Cell 13 pp. 65-71 (1978).
 Taniguchi et al. Gene 10 pp. 11-15 (1980).
 Taniguchi et al. Proc. Japan Acad. 55 (Ser.B) (1979) pp. 464-469.

Goeddel et al. Nature 281 pp. 544-548 (1979).
 Chang et al. Nature 275 pp. 617-624 (1978).
 Derynck et al. Nature 285 pp. 542-547 (1980).

Primary Examiner—Garnette D. Draper
Assistant Examiner—Shelly Guest Cermak

[57]

ABSTRACT

A cDNA library is constructed using mRNA from human fibroblasts induced with poly(I):poly(C). A bacterial clone containing fibroblast interferon cDNA sequences identified by hybridization to a cDNA probe synthesized using deoxyoligonucleotide primers which hybridize to fibroblast interferon mRNA specifically. Expression plasmids are constructed which permit the synthesis in *E. coli* of 8×10^7 units of human fibroblast interferon per liter of culture. The bacterially produced fibroblast interferon is indistinguishable from authentic human fibroblast interferon by several criteria.

6 Claims, 6 Drawing Sheets

| | | | | | |
|----------------------------------|---|----------------------------------|------------------------------|------------------------------|-------------------|
| <u>Protein</u> | 1 | 2 | 3 | 4 | |
| | Met | Ser | Tyr | Asn | - |
| <u>mRNA</u> | | | | | |
| | (5') | AUG-UC ^G _C | UA ^U _C | AA ^U _C | (16 combinations) |
| | (5') | AUG-AG ^U _C | UA ^U _C | AA ^C _U | (8 combinations) |
| <u>Complementary DNA primers</u> | | | | | |
| | ATT- ^A _G TA- ^T _C GA-CAT | | | | Pool 1 |
| | ATT- ^A _G TA- ^A _G GA-CAT | | | | Pool 2 |
| | ATT- ^A _G TA- ^A _G CT-CAT | | | | Pool 3 |
| | GTT- ^A _G TA- ^T _C GA-CAT | | | | Pool 4 |
| | GTT- ^A _G TA- ^A _G GA-CAT | | | | Pool 5 |
| | GTT- ^A _G TA- ^A _G CT-CAT | | | | Pool 6 |

FIG. 1

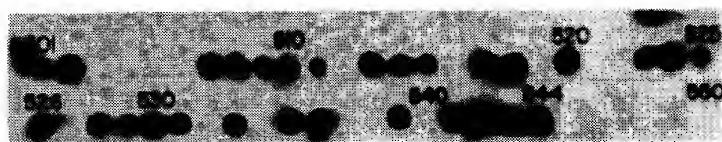


FIG. 2A

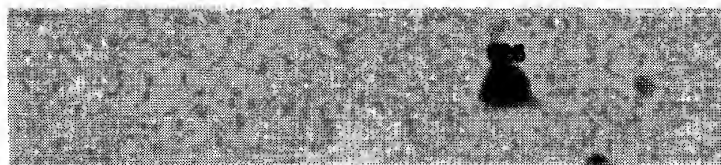


FIG. 2B

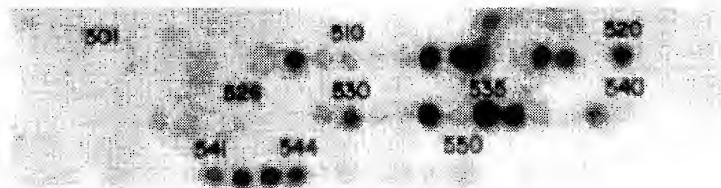


FIG. 2C

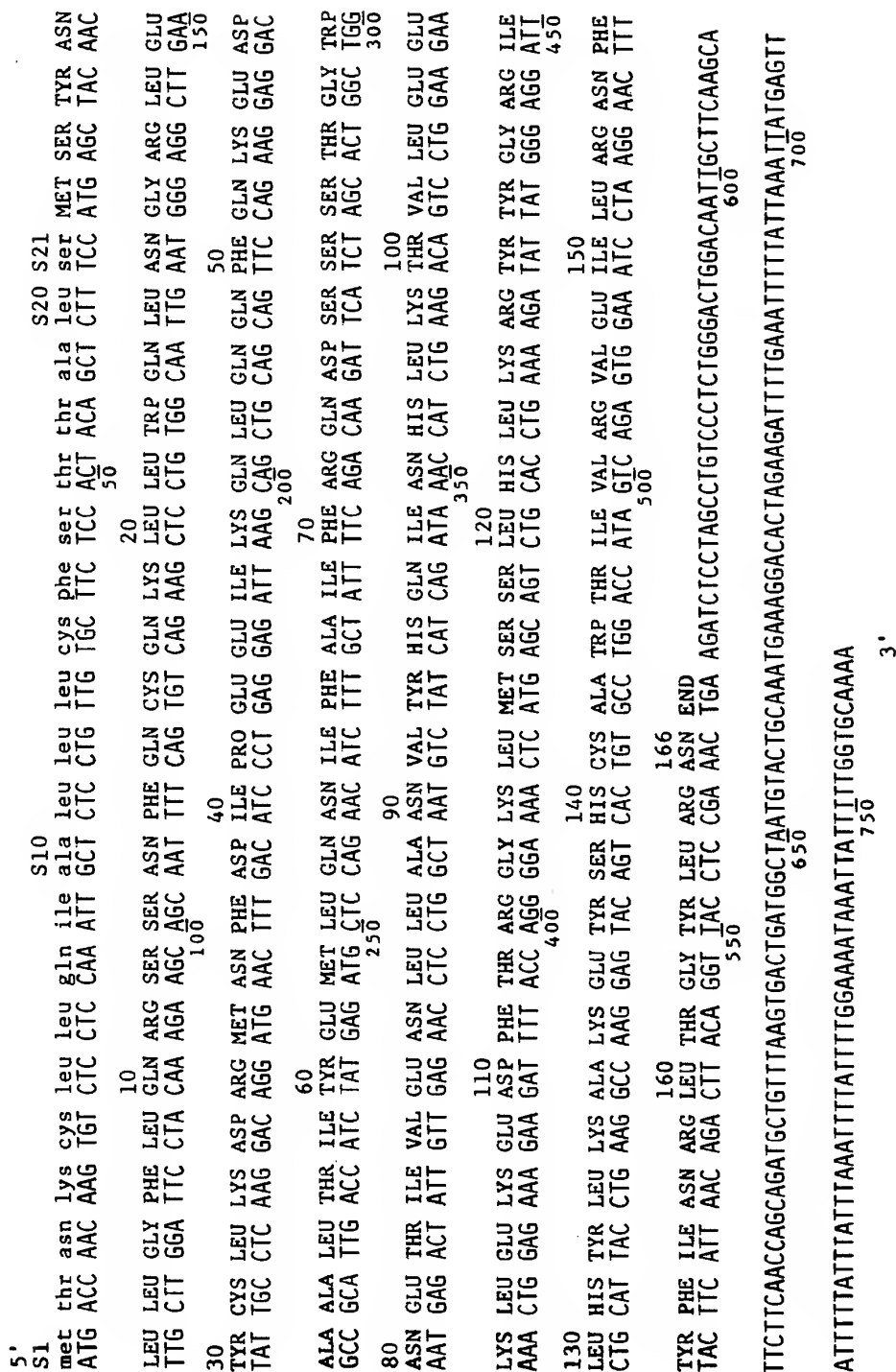


FIG. 3

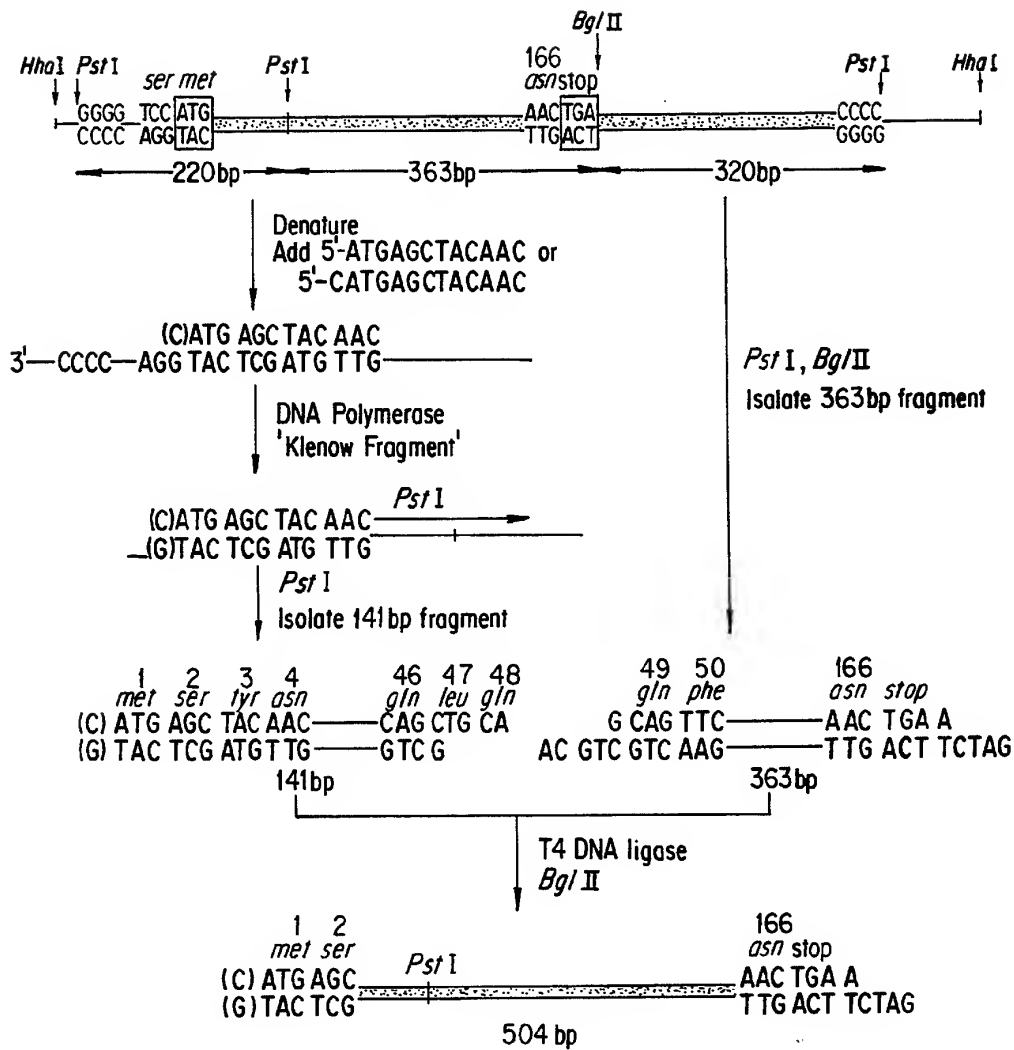


FIG.4

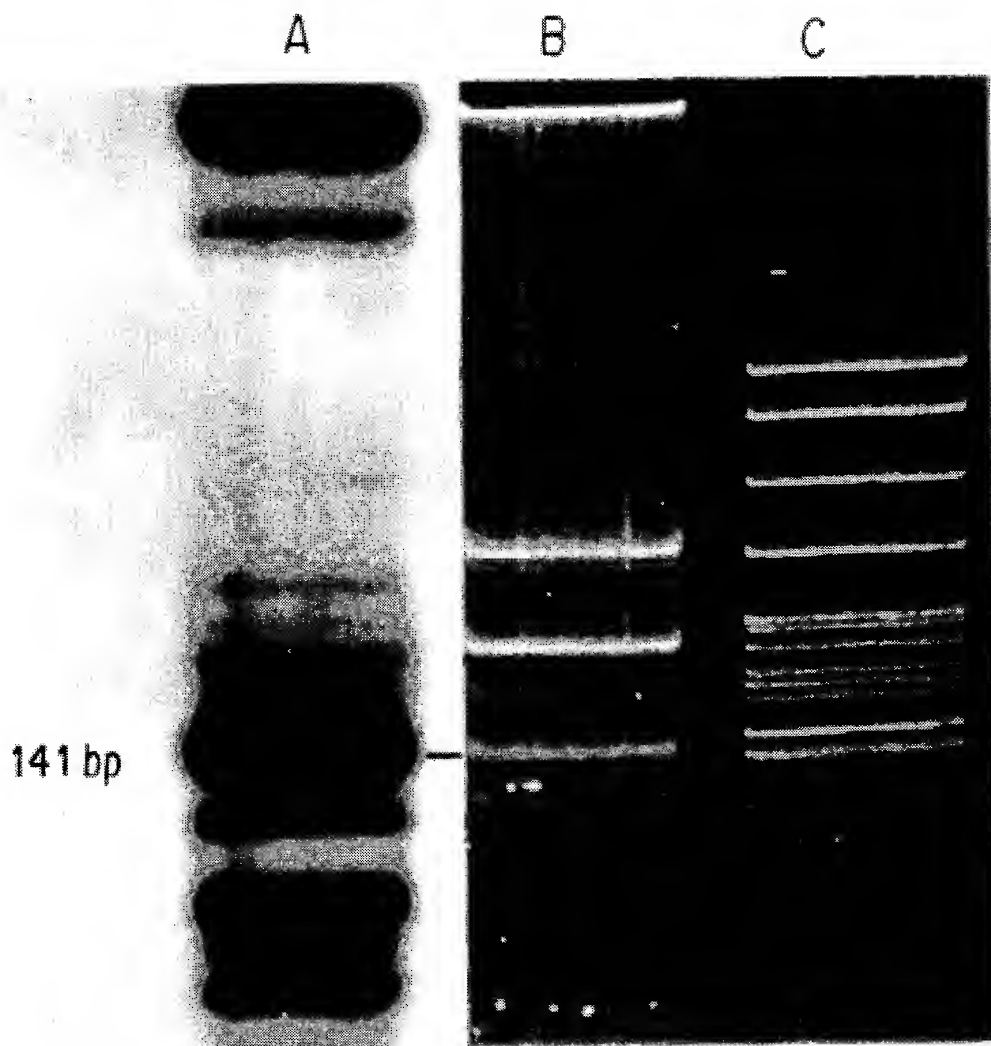


FIG. 5

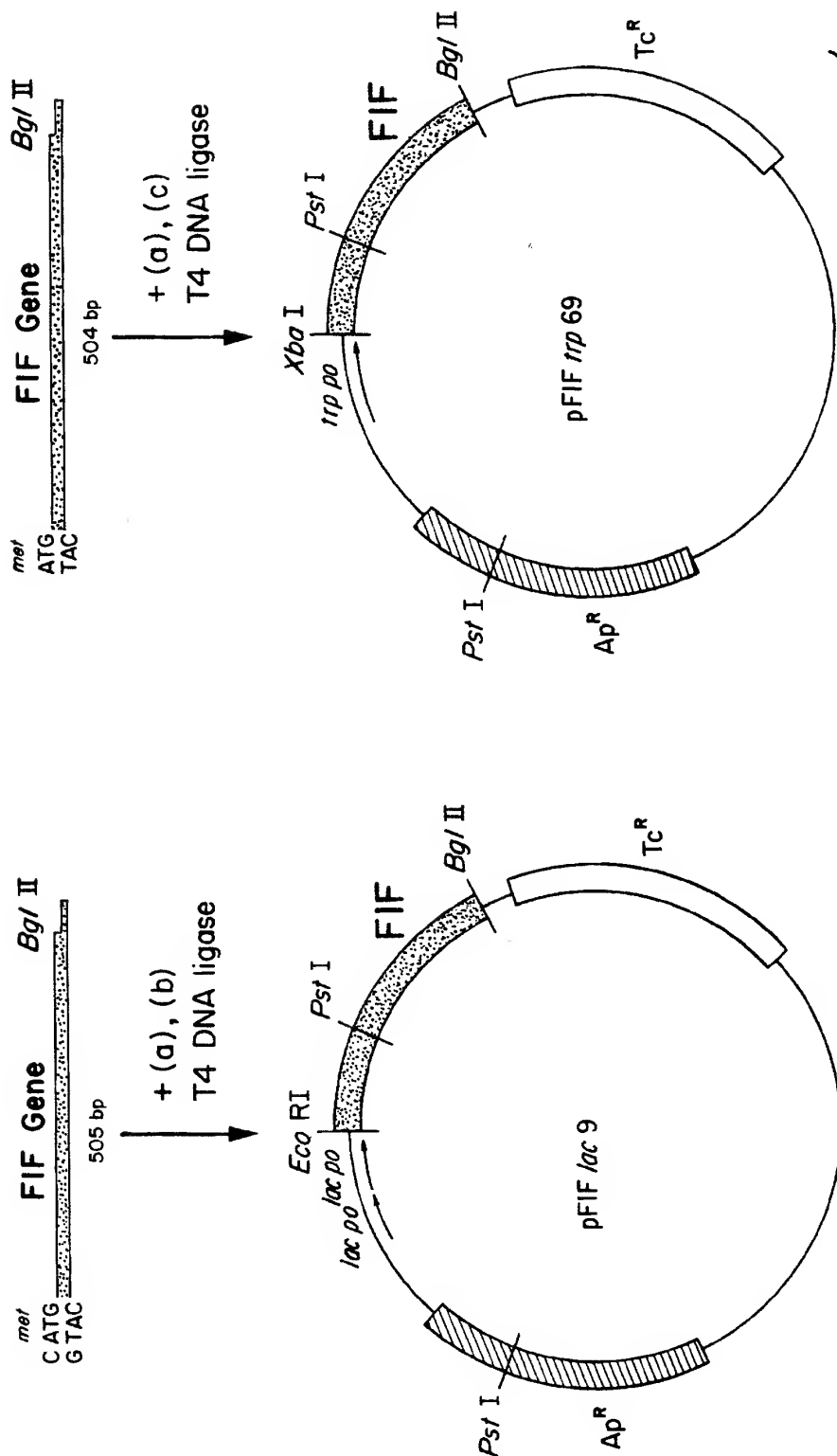


FIG. 6



FIG. 6(a)

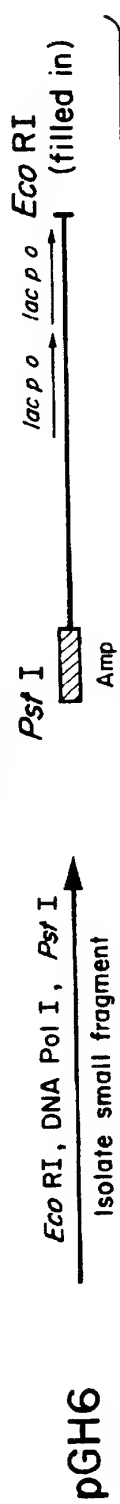


FIG. 6(b)

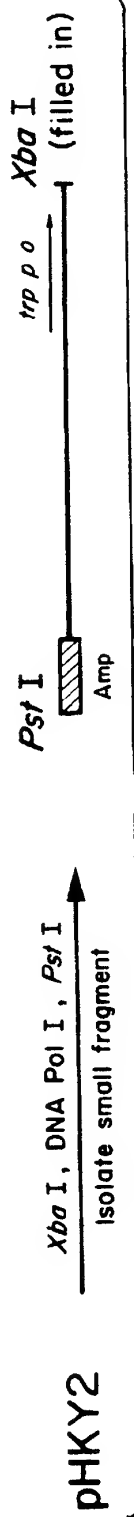


FIG. 6(c)

MATURE HUMAN FIBROBLAST INTERFERON

CROSS-REFERENCE TO RELATED APPLICATION

This is a continuation of application Ser. No. 889,722, filed Jul. 28, 1986, now abandoned which is a divisional of application Ser. No. 291,892, now abandoned filed Aug. 11, 1981, which is a continuation-in-part of application Ser. No. 190,799, filed Sep. 25, 1980, now abandoned.

FIELD OF THE INVENTION

This invention relates to the microbial production, via recombinant DNA technology, of human fibroblast interferon for use in the treatment of viral and neoplastic diseases, and to the means and end products of such production.

BACKGROUND OF THE INVENTION

The publications and other materials referred to herein to illuminate the background of the invention and, in particular cases, to provide additional detail respecting its practice are incorporated herein by reference and, for convenience, are numerically referenced in the following text and respectively grouped in the appended bibliography.

Recombinant DNA Technology

With the advent of recombinant DNA technology, the controlled microbial production of an enormous variety of useful polypeptides has become possible. Already in hand are bacteria modified by this technology to permit the production of such polypeptide products such as somatostatin, the (component) A and B chains of human insulin, human growth hormone. More recently, recombinant DNA techniques have been used to occasion the bacterial production of proinsulin, thymosin alpha 1, (an immune potentiating substance produced by the thymus) and leukocyte interferon.

The workhorse of recombinant DNA technology is the plasmid, a non-chromosomal loop of double-stranded DNA found in bacteria and other microbes, oftentimes in multiple copies per cell. Included in the information encoded in the plasmid DNA is that required to reproduce the plasmid in daughter cells (i.e., a "replicon") and ordinarily, one or more selection characteristics such as, in the case of bacteria, resistance to antibiotics which permit clones of the host cell containing the plasmid of interest to be recognized and preferentially grown in selective media. The utility of plasmids lies in the fact that they can be specifically cleaved by one or another restriction endonuclease or "restriction enzyme", each of which recognizes a different site on the plasmidic DNA. Thereafter heterologous genes or gene fragments may be inserted into the plasmid by endwise joining at the cleavage site or at reconstructed ends adjacent to the cleavage site. DNA recombination is performed outside the cell, but the resulting "recombinant" plasmid can be introduced into it by a process known as transformation and large quantities of the heterologous gene-containing recombinant plasmid obtained by growing the transformant. Moreover, where the gene is properly inserted with reference to portions of the plasmid which govern the transcription and translation of the encoded DNA message, the resulting expression vehicle can be used to actually produce the polypeptide sequence for which the inserted gene codes, a process referred to as expression.

Expression is initiated in a region known as the promoter which is recognized by and bound by RNA polymerase. In some cases, as in the tryptophan or "trp" promoter preferred in the practice of the present invention, promoter regions are overlapped by "operator" regions to form a combined promoter-operator. Operators are DNA sequences which are recognized by so-called repressor proteins which serve to regulate the frequency of transcription initiation at a particular promoter. The polymerase travels along the DNA, transcribing the information contained in the coding strand from its 5' to 3' end into messenger RNA which is in turn translated into a polypeptide having the amino acid sequence for which the DNA codes. Each amino acid is encoded by a nucleotide triplet or "codon" within what may for present purposes be referred to as the "structural gene", i.e. that part which encodes the amino acid sequence of the expressed product. After binding to the promoter, the RNA polymerase first transcribes nucleotides encoding a ribosome binding site, then a translation initiation or "start" signal (ordinarily ATG, which in the resulting messenger RNA becomes AUG), then the nucleotide codons within the structural gene itself. So-called stop codons are transcribed at the end of the structural gene whereafter the polymerase may form an additional sequence of messenger RNA which, because of the presence of the stop signal, will remain untranslated by the ribosomes. Ribosomes bind to the binding site provided on the messenger RNA, in bacteria ordinarily as the mRNA is being formed, and themselves produce the encoded polypeptide, beginning at the translation start signal and ending at the previously mentioned stop signal. The desired product is produced if the sequences encoding the ribosome binding site are positioned properly with respect to the AUG initiator codon and if all remaining codons follow the initiator codon in phase. The resulting product may be obtained by lysing the host cell and recovering the product by appropriate purification from other bacterial protein.

Fibroblast Interferon

Human fibroblast interferon (FIF) is an antiviral protein which also exhibits a wide range of other biological activities (see ref. 1 for review). It has reportedly been purified to homogeneity as a single polypeptide of 19,000-20,000 molecular weight having a specific activity of 2 to 10x10⁸ units/mg (2,3). The sequence of the 13 NH₂-terminal amino acids of FIF has been determined (4). Houghton et al. (5) have used synthetic deoxyoligonucleotides (predicted from this amino acid sequence) to determine the sequence of the 276 5'-terminal nucleotides of FIF mRNA. Taniguchi et al. (6) and Derynck et al. (7) have recently employed RNA selection procedures to identify cloned cDNA copies of FIF mRNA in *E. coli*. See also Taniguchi et al., *Gene* 10, 11 (1980) and *Proc. Natl. Acad. Sci. (U.S.A.)* 77, 5230 (1980) and *Nature* 285, 547 (1980).

While isolation from donor fibroblasts has provided sufficient material for partial characterization and limited clinical studies with homogeneous fibroblast interferon, it is a totally inadequate source for the amounts of interferon heeded for large scale clinical trials and for broad scale prophylactic and/or therapeutic use thereafter. Indeed, presently clinical investigations employing human fibroblast-derived interferon in antitumor and antiviral testing have principally been confined to crude (<1 percent pure) preparations of the material, and long lead times for the manufacture of sufficient quantities, even at unrealistic price levels, have critically delayed investigation on an expanded front.

We perceived that application of recombinant DNA technology would be the most effective way of providing large

quantities of fibroblast interferon which, despite the absence in material so produced of the glycosylation characteristic of human-derived material, could be employed clinically in the treatment of a wide range of viral and neoplastic diseases.

More particularly, we proposed and have since succeeded in producing mature human fibroblast interferon microbially, by constructing a gene therefor which could then be inserted in microbial expression vehicles and expressed under the control of microbial gene regulatory controls.

Our approach to obtaining a fibroblast gene involved the following tasks:

1. Partial amino acid sequences would be obtained by characterization of fibroblast interferon purified to essential homogeneity, and sets of synthetic DNA probes constructed whose codons would, in the aggregate, represent all the possible combinations capable of encoding the partial amino acid sequences.
2. Bacterial colony banks would be prepared containing cDNA from induced messenger RNA. The probes of part (1) would be used to prime the synthesis of radio-labelled single stranded cDNA for use as hybridization probes. The synthetic probes would hybridize with induced mRNA as template and be extended by reverse transcription to form induced, radio-labelled cDNA. Clones from the colony bank that hybridized to radio-labelled cDNA obtained in this manner would be investigated further to confirm the presence of a full-length interferon encoding gene. Any partial length putative gene fragment obtained would itself be used as a probe for the full-length gene.
3. The full-length gene obtained above would be tailored, using synthetic DNA, to eliminate any leader sequence that might prevent microbial expression of the mature polypeptide and to permit appropriate positioning in an expression vehicle relative to start signals and the ribosomal binding site of a microbial promoter. Expressed interferon would be purified to a point permitting confirmation of its character and determination of its activity notwithstanding the absence of glycosylation.

SUMMARY OF INVENTION

A series of replicable plasmidic expression vehicles have been constructed which direct the high level synthesis in transformant microorganisms of a mature polypeptide with the properties of authentic human fibroblast interferon. The product polypeptide exhibits the amino acid sequence of such interferon and is active in *in vitro* testing despite the lack of glycosylation characteristic of the human-derived material. Reference herein to the expression of "mature fibroblast interferon," connotes the bacterial or other microbial production of an interferon molecule unaccompanied by associated glycosylation and the presequence that immediately attends mRNA translation of the human fibroblast interferon genome. Mature fibroblast interferon, according to the present invention, is immediately expressed from a translation start signal (ATG) which also encodes the first amino acid codon of the natural product. The presence or absence of the methionine first amino acid in the microbially expressed product is governed by a kinetic phenomenon dependent on fermentation growth conditions and/or levels of expression in the transformant host. Mature fibroblast interferon could be expressed together with a conjugated protein other than the conventional leader, the conjugate being specifically cleavable in an intra- or extracellular

environment. See British Patent Publication No. 2007676A. Finally, the mature interferon could be produced in conjunction with a microbial "signal" peptide which transports the conjugate to the cell wall, where the signal is processed away and the mature polypeptide secreted.

DESCRIPTION OF THE FIGURES

FIG. 1 depicts the protein sequence information used to design degenerate dodecanucleotide primers. Below the protein sequence are corresponding mRNA sequences and six pools of complementary deoxyoligonucleotide primers.

FIGS. 2A-C are autoradiographs of nitrocellulose-bound plasmid DNA prepared from some of the 600 bacterial transformants having DNA from the fibroblast cDNA library. The plasmid DNA of the nitrocellulose membrane was hybridized with either probe A, B, or C.

FIG. 3 presents DNA sequence of the cDNA insert of clone pFIF3 as determined by the Maxam-Gilbert procedure. The deduced protein sequence is printed above the DNA sequence.

FIG. 4 is a diagram showing the approach used to remove the signal peptide coding regions from pFIF3.

FIG. 5 is an autoradiograph of the polyacrylamide gel used to isolate the 141bp fragment lacking the signal peptide coding regions.

FIG. 6, and 6a, 6b, 6c schematically depicts the construction of plasmids coding for the direct expression of mature fibroblast interferon. Restriction sites and residues are as shown ("Pst I", etc.). "Ap^R" and "Tc^R" connote portions of the plasmid (s) which express, respectively, ampicillin and tetracycline resistance. The legend "p o" is an abbreviation for "promoter operator."

DESCRIPTION OF THE PREFERRED EMBODIMENTS

A. Microorganisms Employed

The work described involved use of the microorganism: *E. coli* K-12 strain 294 (end A, thi⁻, hsr⁻, hsm^k), as described in British Patent Publication No. 2055382 A. This strain has been deposited on Oct. 28, 1978 with the American Type Culture Collection, which is located at 12301 Parklawn Drive, Rockville, Md. 20852, and given ATCC Accession No. 31446. All recombinant DNA work was performed in compliance with applicable guidelines of the National Institutes of Health.

The invention, in its most preferred embodiments, is described with reference to *E. coli*, including not only strain *E. coli* K-12 strain 294, defined above, but also other known *E. coli* strains such as *E. coli* B, *E. coli* x 1776 and *E. coli* W 3110, or other microbial strains many of which are deposited and (potentially) available from recognized microorganism depository institutions, such as the American Type Culture Collection (ATCC)—cf. the ATCC catalogue listing. See also German Offenlegungsschrift 2644432. These other microorganisms include, for example, Bacilli such as *Bacillus subtilis* and other enterobacteriaceae among which can be mentioned as examples *Salmonella typhimurium* and *Serratia marcescens*, utilizing plasmids that can replicate and express heterologous gene sequences therein. Yeast, such as *Saccharomyces cerevisiae*, may also be employed to advantage as host organism in the preparation of the interferon proteins hereof by expression of genes coding therefor under the control of a yeast promoter.

MATERIALS AND METHODS

General methods.

Restriction enzymes were purchased from New England Biolabs and used as directed. Plasmid DNA was prepared by a standard cleared lysate procedure (8) and purified by column chromatography on Biogel A-50M (Bio-Rad). DNA sequencing was performed using the method of Maxam and Gilbert (9). DNA restriction fragments were isolated from polyacrylamide gels by electroelution. DNA fragments were radiolabeled for use as hybridization probes by the random calf thymus DNA priming procedure of Taylor et al. (10). In situ colony hybridizations were performed by the Grunstein-Hogness procedure (11).

Chemical synthesis of deoxyoligonucleotides.

The deoxyoligonucleotides were synthesized by the modified phosphotriester method in solution (12), using trideoxynucleotides as building block (13). The material and general procedures were similar to those described (14). The six pools of primers (Fib 1-6) containing four dodecanucleotides each were obtained by separately coupling two hexamer pools (of two different 5'-terminal sequences each) with three different hexamer pools (of two different 3'-terminal sequences each).

Induction of fibroblasts.

Human fibroblasts (cell line GM-2504A) were grown as described previously (15). Growth medium (Eagles's minimal essential medium containing 10 percent fetal calf serum) was removed from roller bottles (Corning, 850 cm²) and replaced with 50 ml growth medium containing 50 µg/ml of poly (I):poly (C) (PL Biochemicals) and 10 µg/ml cycloheximide. This induction medium was removed after 4 hours at 37° C. and cell monolayers were washed with PBS (0.14M NaCl, 3 mM KCl, 1.5 mM KH₂PO₄, 8 mM Na₂HPO₄). Each bottle was incubated at 37° C. with 10 ml of a trypsin-EDTA solution (Gibco 610-5305) until cells were detached, and fetal calf serum was added to a concentration of 10 percent. Cells were spun for 15 minutes at 500× g and pellets were resuspended in PBS, pooled, and resedimented. Cells were frozen in liquid nitrogen. Approximately 0.17 g of cells were obtained per roller bottle.

Preparation and assay of interferon mRNA.

Poly extraction and oligo(dT)-cellulose chromatography as described elsewhere (16). The poly (A) containing RNA was enriched for interferon mRNA by centrifugation on a linear 5 percent to 20 percent (w/v) sucrose gradient. The RNA samples were heated to 80° C. for 2 minutes, rapidly cooled, layered over the gradient, and centrifuged for 20 hours at 30,000 rpm at 4° C. in a Beckman SW-40 rotor. Fractions were collected, ethanol precipitated, and dissolved in H₂O.

One microgram samples of mRNA were injected into *Xenopus laevis* oocytes as described previously (17,18). The injected oocytes were incubated 24 hours at 21° C., homogenized, and centrifuged for 5 minutes at 10,000× g. The interferon in the supernatant was determined by the cytopathic effect (CPE) inhibition assay (1) using Sindbis virus and human diploid (WISH) cells. Interferon titers of 1,000 to 6,000 units recovered (NIH reference standard) per microgram of RNA injected were routinely obtained for the 12S species of mRNA.

Synthesis and cloning of cDNA.

Single stranded cDNA was prepared in 100 µl reactions containing 5 µg of 12S fraction mRNA, 20 mM Tris-HCl (pH 8.3), 20 mM KCl, 8 mM MgCl₂, 30 mM β-mercaptoethanol, 100 µCi of (α³²P)dCTP (Amersham) and 1 mM dATP, dCTP, dGTP, dTTP. The primer was the synthetic Hind III decamer dCCAAGCTTGG (19), which had been extended at the 3' terminus with about 20 to 30 deoxythymidine residues using terminal deoxynucleotidyl transferase

(20). 100 units of AMV reverse transcriptase were added and the reaction mixture was incubated at 42° C. for 30 minutes. The second strand DNA synthesis was carried out as described previously (21). The double stranded cDNA was treated with 1200 units of S1 nuclease (Miles Laboratories) for 2 hours at 37° C. in 25 mM sodium acetate (pH 4.5), 1 mM ZnCl₂, 0.3M NaCl. After phenol extraction the mixture was separated electrophoretically on an 8 percent polyacrylamide gel. cDNA (~0.5 µg) ranging from 500 to 1500 base pairs in size was recovered by electroelution. A 20 ng aliquot was extended with deoxyC residues using terminal deoxynucleotidyl transferase (20), and annealed with 100 ng of pBR322 which had been cleaved with Pst I and tailed with deoxyG residues (20). The annealed mixture was used to transform *E. coli* k-12 strain 294 (22) by a published procedure (23). Strain 294 was used throughout in the work described here, and has been deposited with the American Type Culture collection, accession no. 31446.

Preparation of induced and uninduced ³²p-cDNA probes. 5 µg of 12S mRNA were combined with either 2 µg of oligo (dT)₁₂₋₁₈ (Collaborative Research) or 5 µg of each synthetic primer pool (Fib 1 to Fib 6) in 60 µl of 10 mM Tris-HCl (pH 8), 1 mM EDTA. The mixtures were boiled 3 minutes, and quenched on ice. 60 µl of 40 mM Tris-HCl (pH 8.3), 40 mM KCl, 16 mM MgCl₂, 60 mM β-mercaptoethanol, 1 mM dATP, dGTP, dTTP and 5×10⁻⁷M (α-³²P) dCTP (Amersham, 2,000-3,000 Ci/mmol) was added to each template-primer mix at 0° C. After the addition of 100 units of AMV reverse transcriptase, the reactions were incubated at 42° C. for 30 minutes and purified by passage over 10 ml Sephadex G-50 columns. The products were treated with 0.3N NaOH for 30 minutes at 70° C., neutralized, and ethanol precipitated.

The ³²p-cDNAs were combined with 100 µg of poly (A) mRNA from uninduced fibroblasts in 50 µl of 0.4M sodium phosphate (pH 6.8), 0.1 percent SDS. The mixtures were heated at 98° C. for 5 minutes and allowed to anneal 15 hours at 45° C. The DNA-RNA hybrids, (containing uninduced cDNA sequences) were separated from single-stranded DNA (induced cDNA sequences) by chromatography on hydroxyapatite as described by Galau et al. (24). The DNA-RNA hybrids were treated with alkali to remove RNA.

Screening of recombinant plasmids with ³²p-cDNA probes.

Approximately 1 µg samples of plasmid DNA were prepared from individual transformants by a published procedure (25). The DNA samples were linearized by digestion with Eco RI, denatured in alkali, and applied to each of three nitrocellulose filters (Schleicher and Schuell, BA85) by the dot hybridization procedure (26). The filters were hybridized with the ³²p-cDNA probes for 16 hours at 42° C. in 50 percent formamide, 10× Denhardt's solution (27), 6×SSC, 40 mM Tris-HCl (pH 7.5), 2 mM EDTA, 40 µg/ml yeast RNA. Filters were washed with 0.1×SSC, 0.1 percent SDS twice for 30' at 42° C., dried, and exposed to Kodak XR-2 x-ray film using Dupont Lightning-Plus intensifying screens at -80° C.

Construction of plasmids for direct expression of FIF.

The synthetic primers I (dATGAGCTACAAC) and II (dCATGAGCTACAAC) were phosphorylated using T4 polynucleotide kinase and (γ-³²P)ATP (Amersham) to a specific activity of 700 Ci/mmol as described previously (28). Primer repair reactions were performed as follows: 250 pmoles of the ³²P-primers were combined with 8 µg (10 pmole) of a 1200 bp Hha I restriction fragment containing the FIF cDNA sequence. The mixture was ethanol precipitated, resuspended in 50 µl H₂O, boiled 3 minutes, quenched in a dry ice-ethanol bath, and combined with a 50 µl solution

of 20 mM Tris-HCl (pH 7.5), 14 mM MgCl₂, 120 mM NaCl, 0.5 mM dATP, dCTP, dGTP, dTTP at 0° C. 10 units of DNA polymerase I Klenow fragment (Boehringer-Mannheim) were added and the mixture was incubated at 37° C. for 4½ hours. Following extraction with phenol/CHCl₃ and restriction with Pst I, the desired product was purified on a 6 percent polyacrylamide gel. Subsequent ligations were done at room temperature (cohesive termini) or 4° C. (blunt ends) using previously detailed conditions (21,28).

Assay for interferon expression in *E. coli*.

Bacterial extracts were prepared for IF assay as follows: One ml cultures were grown overnight in LB (29) containing 5 µg/ml tetracycline, then diluted into 25 ml of M9 medium (29) containing 0.2 percent glucose, 0.5 percent casamino acids and 5 µg/ml tetracycline. 10 ml samples were harvested by centrifugation when A₅₅₀ (Absorbance at 500 nanometers) reached 1.0. The cell pellets were quickly frozen in a dry ice-ethanol bath and cleared lysates were prepared as described by Clewell (8). Interferon activity in the supernatants was determined by comparison with NIH FIF standards using cytopathic effect (CPE) inhibition assays as reviewed previously (1). Two different assays were used: (a) WISH (human amnion) cells were seeded in microtiter dishes. Samples were added 16 to 20 hours later and diluted by serial 2-fold dilution. Sindbis virus was added after at least 3 hours in incubation. Plates were stained 20 to 24 hours later with crystal violet. (b) MDBK (bovine kidney) cell line was seeded simultaneously with 2-fold dilutions of samples. Vesicular stomatitis virus was added after 2 to 3 hours incubation and plates were stained with crystal violet 16 to 18 hours later. To test pH 2 stability bacterial extracts and standards were diluted in minimal essential medium to a concentration of 1000 units/ml. One ml aliquots were adjusted to pH 2 with 1N HCl, incubated at 4° C. for 16 hours, and neutralized by addition of NaOH. IF activity was determined by the CPE inhibition assay using human amnion cells. To establish antigenic identity 25 µl aliquots of the 1000 U/ml interferon samples (untreated) were incubated with 25 µl of rabbit antihuman leukocyte interferon for 60' at 37° C., centrifuged at 12,000× g for 5 minutes and the supernatant assayed. Fibroblast and leukocyte interferon standards were obtained from the National Institutes of Health. Rabbit antihuman leukocyte interferon was obtained from the National Institute of Allergy and Infectious Diseases.

RESULTS

Chemical synthesis of primer pools complementary of FIF mRNA.

The amino-terminal protein sequence of human fibroblast interferon (4) permitted us to deduce the 24 possible mRNA sequences which could code for the first four amino acids. The 24 complementary deoxyoligonucleotides were synthesized in 6 pools of 4 dodecamers each (FIG. 1).

The six pools of 4 deoxyoligonucleotides each were synthesized by a modified phosphotriester method that has been used previously for the rapid synthesis of oligonucleotides in solution (12) and on solid phase (14). The basic strategy involved reacting two different 3'-blocked trimers with an excess of a single 5'-protected trimer to yield a pool of two hexamers, each represented equally. The coupling of two pools, each containing two hexamers, then resulted in a pool of four dodecamers.

Identification of FIF cDNA clones.

Using 12S mRNA from induced human fibroblasts (1,000 units IF activity per µg in oocyte assay), double stranded cDNA was prepared and inserted into pBR322 at the Pst I site by the standard dG:dC tailing method (20). A fibroblast

cDNA library consisting of 30,000 ampicillin-sensitive, tetracycline-resistant transformants of *E. coli* K-12 strain 294 was obtained from 20 ng of cDNA ranging in size from 550 to 1300 base pairs. Plasmid DNA was prepared from 600 of the transformants and applied to 3 sets of nitrocellulose filters as described in Materials and Methods.

The approach followed in the identification of hybrid plasmids containing fibroblast interferon cDNA sequences was similar to that used to identify human leukocyte interferon recombinant plasmids (30). Radiolabeled cDNA hybridization probes were prepared using either the 24 synthetic dodecamers or oligo(dT)₁₂₋₁₈ as primers and 12S RNA from induced fibroblasts (5000 units/µg in oocytes) as template. The ³²P-cDNAs (specific activity >5×10⁸ cpm/µg) obtained were hybridized to a large excess of mRNA isolated from uninduced human fibroblasts, and the mRNA-cDNA hybrids were separated from unreacted cDNA by hydroxyapatite chromatography (24). The single stranded cDNA fractions should be enriched for sequences which are present in induced fibroblasts but absent in uninduced cells, and the mRNA-cDNA hybrids should represent sequences common to both induced and uninduced cells. Approximately 4×10⁶ cpm of single stranded cDNA (hybridization probe A) and 8×10⁶ cpm of cDNA-mRNA hybrids were obtained using oligo(dT)₁₂₋₁₈ primed cDNA; 1.5×10⁶ cpm of single stranded (hybridization probe B) and 1.5×10⁵ cpm of hybrids were obtained from cDNA primed using synthetic dodecamer pools Fib 1-6. The cDNA-mRNA hybrids from both fractionations were combined, the RNA hydrolyzed by treatment with alkali, and the ³²P-cDNA used as hybridization probe C. Many of the 600 plasmid samples hybridized with both probes A and C, indicating that the hybridization reactions between uninduced mRNA and ³²P-cDNA (prior to the hydroxyapatite fractionation step) had not gone to completion. However, only one of the 600 plasmids (pF526) hybridized strongly with the specifically primed, induced cDNA probe B (FIG. 2). Plasmid pF526 also hybridized with the total oligo(dT)₁₂₋₁₈ primed, induced cDNA probe A, and failed to give detectable hybridization to the combined uninduced probe C.

Pst I digestion of pF526 showed the cloned cDNA insert to be about 550 base pairs long, probably too short to contain the entire coding region for a protein the size of fibroblast interferon. Therefore, a ³²P-labeled DNA probe was prepared from this Pst I fragment by random priming with calf thymus DNA (10). This probe was used to screen 2000 individual colonies from a newly constructed fibroblast cDNA library (the new cDNA library was prepared using 12S mRNA from induced fibroblasts having a titer of 6,000 units/ml in the oocyte assay system). Sixteen clones hybridized to the probe. Plasmids prepared from the majority of these released two fragments when cleaved with Pst I, indicating that the cDNA contained an internal Pst I site. Clone pFIF3 contained the largest cDNA insert, about 800 base pairs. The DNA sequence of the insert was determined by the Maxam-Gilbert procedure (9) and is shown in FIG. 3. The amino acid sequence of human fibroblast interferon predicted from the nucleotide sequence is identical to that reported recently by Taniguchi et al. (31) and by Derynck et al. (7) from DNA sequencing of FIF cDNA clones. A precursor or signal peptide of 21 amino acids is followed by a mature interferon polypeptide of 166 amino acids, a stretch of 196 3'-untranslated nucleotides and a poly(A) tail. The NH₂-terminal 20 amino acids of mature FIF have been directly determined by protein microsequencing and are the same as those predicted from the DNA sequence. The calculated formula molecular weight of mature human fibro-

blast interferon having the 166 amino acids shown in FIG. 3 is about 20,027.

Direct expression of fibroblast interferon.

To express high levels of mature fibroblast interferon in *E. coli*, initiation of protein synthesis must occur at the ATG codon of the mature polypeptide (amino acid 1) rather than at the ATG of the signal peptide (amino acid S1) (FIG. 3).

Our approach to removing the signal peptide coding regions from pFIF3 is depicted in FIG. 4. A 1200 bp DNA fragment which contained the entire cDNA insert was isolated from a polyacrylamide gel after digesting pFIF3 with Hha I. Two separate synthetic deoxyligonucleotide primers, dATGAGCTACAAC(I) and dCATGAGCTACAAC(II), were prepared. Both primers contain the coding sequence for the first four amino acids of mature fibroblast interferon; primer II has an additional C at the 5'-terminus. Primer repair reactions and subsequent ligations were carried out separately for primers I and II, and gave nearly identical results. Therefore, only reactions using primer I are discussed in detail here. The primers were 5'-radiolabeled using (γ -³²P)ATP and T4 polynucleotide kinase, combined with the 1200 bp Hha I DNA fragment and the mixture denatured by boiling. Following hybridization of the primer to the denatured Hha I DNA fragment, *E. coli* DNA polymerase I Klenow fragment (33) was used to catalyze the repair synthesis of the plus (top) strand (FIG. 4). In addition, the associated 3' \rightarrow 5' exonuclease activity of the Klenow fragment removed the 3'-protruding end from the minus (bottom) strand, leaving a flush end. Analysis of samples of the reaction mixture by polyacrylamide gel electrophoresis indicated that the repair synthesis did not go to completion, but stopped at several discrete sites. Therefore, the entire reaction mixture was treated with Pst I and the desired 141 bp fragment (180,000 Cerenkov cpm; \sim 0.3 pmole) was purified by polyacrylamide gel electrophoresis (FIG. 5). Ligation of this fragment to 1 μ g (\sim 4 pmole) of the 363 bp Pst I-Bgl II fragment isolated from pFIF3 (FIG. 4), followed by Bgl II digestion, yielded 50,000 Cerenkov cpm (\sim 0.1 pmole, \sim 30 ng) of the 504 bp DNA fragment containing the entire coding sequence for mature fibroblast interferon. The same reactions using primer II gave 83,000 cpm (\sim 0.15 pmole, \sim 50 ng) of 505 bp product.

The construction of plasmids which direct the synthesis of human fibroblast interferon is outlined in FIG. 6. Separate expression plasmids were constructed which placed FIF synthesis under the control of the *E. coli* lac or trp promoter-operator systems. Both of these systems have proven useful for the direct expression of eukaryotic genes in *E. coli*: human growth hormone has been efficiently synthesized using the lac system (21) and human leukocyte interferon has been produced at high levels using the trp system (30) and *Nature* 287, 411 (1980).

pBRH trp was digested with EcoRI restriction enzyme and the resulting fragment isolated by PAGE and electroelution. EcoRI-digested plasmid pSom 11 (K. Itakura et al., *Science* 198, 1056 (1977); G.B. patent publication no. 2,007,676 A) was combined with the above fragment. The mixture was ligated with T₄ DNA ligase as previously described and the resulting DNA transformed into *E. coli* K-12 strain 294 as previously described. Transformant bacteria were selected on ampicillin-containing plates. Resulting ampicillin-resistant colonies were screened by colony hybridization (M. Gruenstein et al., *Proc Nat'l Acad Sci USA* 72, 3951-3965 [1975]) using as a probe the trp promoter-operator-containing the above fragment isolated from pBRHtrp, which had been radioactively labeled with p³². Several colonies shown positive by colony hybridization

were selected, plasmid DNA was isolated and the orientation of the inserted fragments determined by restriction analysis employing restriction enzymes BglII and BamHI in double digestion. *E. coli* 294 containing the plasmid designated pSOM7A2, which has the trp promoter-operator fragment in the desired orientation was grown in LB medium containing 10 μ g/ml ampicillin. The cells were grown to optical density 1 (at 550 nm), collected by centrifugation and resuspended in M9 media in tenfold dilution. Cells were grown for 2-3 hours, again to optical density 1, then lysed and total cellular protein analyzed by SDS (sodium dodecyl sulfate) urea (15 percent) polyacrylamide gel electrophoresis (J. V. Maizel Jr. et al., *Meth Viral* 5, 180-246 [1971]).

Plasmid pBR322 was Hind III digested and the protruding Hind III ends in turn digested with S1 nuclease. The S1 nuclease digestion involved treatment of 10 μ g of Hind III-cleaved pBR322 in 30 μ l S1 buffer (0.3M NaCl, 1 mM ZnCl₂, 25 mM sodium acetate, pH 4.5) with 300 units S1 nuclease for 30 minutes at 15° C. The reaction was stopped by the addition of 1 μ l of 30 \times S1 nuclease stop solution (0.8M tris base, 50 mM EDTA). The mixture was phenol extracted, chloroform extracted and ethanol precipitated, then EcoRI digested as previously described and the large fragment (1) obtained by PAGE procedure followed by electroelution. The fragment obtained has a first EcoRI sticky end and a second, blunt end whose coding strand begins with the nucleotide thymidine.

Plasmid pSom7 A2, as prepared above, was Bgl II digested and the Bgl II sticky ends resulting made double stranded with the Klenow polymerase I procedure using all four deoxynucleotide triphosphates. EcoRI cleavage of the resulting product followed by PAGE and electroelution of the small fragment (2) yielded a linear piece of DNA containing the tryptophan promoter-operator and codons of the LE' "Proximal" sequence upstream from the Bgl II site ("LE'(p)"). The product had an EcoRI end and a blunt end resulting from filling in the Bgl II site. However, the Bgl II site is reconstituted by ligation of the blunt end of the above fragment (2) to the blunt end of the above prepared fragment (1). Thus, the two fragments were ligated in the presence of T₄ DNA ligase to form the recirculated plasmid pHKY 10 which was propagated by transformation into competent *E. coli* strain 294 cells.

Plasmid pGM1 carries the *E. coli* tryptophan operon containing the deletion ALE1413 (G. F. Miozzari, et al., (1978) *J. Bacteriology* 133, 1457-1466) and hence expresses a fusion protein comprising the first 6 amino acids of the trp leader and approximately the last third of the trp E polypeptide (hereinafter referred to in conjunction as LE'), as well as the trp D polypeptide in its entirety, all under the control of the trp promoter-operator system. The plasmid, 20 μ g, was digested with the restriction enzyme PvuII which cleaves the plasmid at five sites. The gene fragments were next combined with EcoRI linkers (consisting of a self complementary oligonucleotide of the sequence: pCAT-GAATTCATG) providing an EcoRI cleavage site for a later cloning into a plasmid containing an EcoRI site. The 20 μ g of DNA fragments obtained from pGM1 were treated with 10 units T₄ DNA ligase in the presence of 200 pico moles of the 5'-phosphorylated synthetic oligonucleotide pCAT-GAATCATG and in 20 μ l T₄ DNA ligase buffer (20 mM tris, pH 7.6, 0.5 mM ATP, 10 mM MgCl₂, 5 mM dithiothreitol) at 4° C. overnight. The solution was then heated 10 minutes at 70° C. to halt ligation. The linkers were cleaved by EcoRI digestion and the fragments, now with EcoRI ends were separated using 5 percent polyacrylamide gel electrophoresis (hereinafter "PAGE") and the three largest fragments

isolated from the gel by first staining with ethidium bromide, locating the fragments with ultraviolet light, and cutting from the gel the portions of interest. Each gel fragment, with 300 microliters 0.1×TBE, was placed in a dialysis bag and subjected to electrophoresis at 100 v for one hour in 0.1× TBE buffer (TBE buffer contains: 10.8 gm tris base, 5.5 gm boric acid, 0.09 gm Na₂EDTA in 1 liter H₂O). The aqueous solution was collected from the dialysis bag, phenol extracted, chloroform extracted and made 0.2M sodium chloride, and the DNA recovered in water after ethanol precipitation. The trp promoter-operator-containing gene with EcoRI sticky ends was identified in the procedure next described, which entails the insertion of fragments into a tetracycline sensitive plasmid which, upon promoter-operator insertion, becomes tetracycline resistant.

Plasmid pBRH1 (R. I. Rodriguez, et al, Nuclcic Acids Research 6, 3267-3287 [1979]) expresses ampicillin resistance and contains the gene for tetracycline resistance but, there being no associated promoter, does not express that resistance. The plasmid is accordingly tetracycline sensitive. By introducing a promoter-operator system in the EcoRI site, the plasmid can be made tetracycline resistant.

pBRH1 was digested with EcoRI and the enzyme removed by phenol extraction followed by chloroform extraction and recovered in water after ethanol precipitation. The resulting DNA molecule was, in separate reaction mixtures, combined with each of the three DNA fragments obtained above and ligated with T₄ DNA ligase as previously described. The DNA present in the reaction mixture was used to transform competent *E. coli* K-12 strain 294, K. Backman et al., Proc Nat'l Acad Sci USA 73, 4174-4198 [1976]) by standard techniques (V. Hershsfield et al., Proc Nat'l Acad Sci USA 71, 3455-3459 [1974]) and the bacteria plated on LB plates containing 20 µg/ml ampicillin and 5 µg/ml tetracycline. Several tetracycline-resistant colonies were selected, plasmid DNA isolated and the presence of the desired fragment confirmed by restriction enzyme analysis. The resulting plasmid is designated pBRHtrp.

An EcoRI and BamHI digestion product of the viral genome of hepatitis B was obtained by conventional means and cloned into the EcoRI and BamHI sites of plasmid pGII6 (D. V. Goeddel et al., Nature 281, 544 [1979]) to form the plasmid pHS32. Plasmid pHS32 was cleaved with XbaI, phenol extracted, chloroform extracted and ethanol precipitated. It was then treated with 1 µl *E. coli* polymerase I, Klenow fragment (Boehringer-Mannheim) in 30 µl polymerase buffer (50 mM potassium phosphate pH 7.4, 7 mM MgCl₂, 1 mM β-mercaptoethanol) containing 0.1 mM dTTP and 0.1 mM dCTP for 30 minutes at 0° C. then 2 hr. at 37° C. This treatment causes 2 of the 4 nucleotides complementary to the 5' protruding end of the XbaI cleavage site to be filled in: t,0170

Two nucleotides, dC and dT, were incorporated giving an end with two 5' protruding nucleotides. This linear residue of plasmid pHS32 (after phenol and chloroform extraction and recovery in water after ethanol precipitation) was cleaved with EcoRI. The large plasmid fragment was separated from the smaller EcoRI-XbaI fragment by PAGE and isolated after electroelution. This DNA fragment from pHS32 (0.2 µg), was ligated, under conditions similar to those described above, to the EcoRI-Taq I fragment of the tryptophan operon (~0.01 µg), derived from pBRHtrp.

In the process of ligating the fragment from pHS32 to the Eco RI-Taq I fragment, as described above, the Taq I protruding end is ligated to the XbaI remaining protruding end even though it is not completely Watson-Crick base-paired: t,0180

A portion of this ligation reaction mixture was transformed into *E. coli* 294 cells, heat treated and plated on LB plates containing ampicillin. Twenty-four colonies were selected, grown in 3 ml LB media, and plasmid isolated. Six of these were found to have XbaI site regenerated via *E. coli* catalyzed DNA repair and replication: t,0181

These plasmids were also found to cleave both with EcoRI and HpaI and to give the expected restriction fragments. One plasmid, designated pTrp 14, was used for expression of heterologous polypeptides, as next discussed.

The plasmid pHGH 107 (D. V. Goeddel et al, Nature, 281, 544, 1979) contains a gene for human growth hormone made up of 23 amino acid codons produced from synthetic DNA fragments and 163 amino acid codons obtained from complementary DNA produced via reverse transcription of human growth hormone messenger RNA. This gene, though it lacks the codons of the "pre" sequence of human growth hormone, does contain an ATG translation initiation codon. The gene was isolated from 10 µg pHGH 107 after treatment with EcoRI followed by *E. coli* polymerase I Klenow fragment and dTTP and dATP as described above. Following phenol and chloroform extraction and ethanol precipitation the plasmid was treated with BamHI.

The human growth hormone ("HGH") gene-containing fragment was isolated by PAGE followed by electroelution. The resulting DNA fragment also contains the first 350 nucleotides of the tetracycline resistance structural gene, but lacks the tetracycline promoter-operator system so that, when subsequently cloned into an expression plasmid, plasmids containing the insert can be located by the restoration of tetracycline resistance. Because the EcoRI end of the fragment has been filled in by the Klenow polymerase I procedure, the fragment has one blunt and one sticky end, ensuring proper orientation when later inserted into an expression plasmid.

The expression plasmid pTrp14 was next prepared to receive the HGH gene-containing fragment prepared above. Thus, pTrp14 was XbaI digest and the resulting sticky ends filled in with the Klenow polymerase I procedure employing dATP, dTTP, dGTP and dCTP. After phenol and chloroform extraction and ethanol precipitation the resulting DNA was treated with BamHI and the resulting large plasmid fragment isolated by PAGE and electroelution. The pTrp14-derived fragment had one blunt and one sticky end, permitting recombination in proper orientation with the HGH gene containing fragment previously described.

The HGH gene fragment and the pTrp14 ΔXba-BamHI fragment were combined and ligated together under conditions similar to those described above. The filled in XbaI and EcoRI ends ligated together by blunt end ligation to recreate both the XbaI and the EcoRI site: t,0190

This construction also recreates the tetracycline resistance gene. Since the plasmid pHGH 107 expresses tetracycline resistance from a promoter lying upstream from the HGH gene (the lac promoter), this construction, designated pHGH 207, permits expression of the gene for tetracycline resistance under the control of the tryptophan promoter-operator. Thus the ligation mixture was transformed into *E. coli* 294 and colonies selected on LB plates containing 5 µg/ml tetracycline.

Plasmid pHGH 207 was EcoRI digested and the trp promoter containing EcoRI fragment recovered by PAGE followed by electroelution. Plasmid pBRH1 was EcoRI digested and the cleaved ends treated with bacterial alkaline phosphatase ("BAP") (1 µg, in 50 mM tris pH 8 and 10 mM MgCl₂ for 30 min. at 65° C.) to remove the phosphate groups on the protruding EcoRI ends. Excess bacterial

alkaline phosphatase was removed by phenol extraction, chloroform extraction and ethanol precipitation. The resulting linear DNA, because it lacks phosphates on the protruding ends thereof, will in ligation accept only inserts whose complementary stick ends are phosphorylated but will not itself recircularize, permitting more facile screening for plasmids containing the inserts.

The EcoRI fragment derived from pHGH 207 and the linear DNA obtained from pBRH1 were combined in the presence of T₄ ligase as previously described and ligated. A portion of the resulting mixture was transformed into *E. coli* strain 294 as previously described, plated on LB media containing 5 µg/ml of tetracycline, and 12 tetracycline resistant colonies selected. Plasmid was isolated from each colony and examined for the presence of a DNA insert by restriction endonuclease analysis employing EcoRI and XbaI. One plasmid containing the insert was designated pHKY1.

The plasmid pHKY10, described above, is a derivative of pBR322 which contains a Bgl II site between the tetracycline resistance (Tc^R) promoter and structural gene (32). The large DNA fragment isolated after digesting pHKY10 with Pst I and Bgl II therefore contains part of the ampicillin resistance (Ap^R) gene and all of the Tc^R structural gene, but lacks the Tc^R promoter (FIG. 6). The plasmid pGH6 (21) was digested with Eco RI, the resulting single stranded ends were filled in with DNA polymerase I, and the plasmid was cleaved with Pst I. The small fragment, containing part of the Ap^R gene, a double lac promoter and lac ribosome binding site, but lacking an ATG initiation triplet was isolated. A similar trp promoter fragment, containing the trp leader ribosome binding site, but lacking an ATG sequence (30), may be isolated from pHKY1 described above; see (32) (see FIG. 6).

The trp fragment just referred to is an analog of the *E. coli* tryptophan operon from which the so-called trp attenuator has been deleted, See *J. Bact.* 133, 1457 (1978), to controllably heighten expression levels. Expression plasmids containing the modified trp regulon can be grown to predetermined levels in nutrient media containing additive tryptophan in quantities sufficient to repress the promoter-operator system, then be deprived of tryptophan so as to derepress the system and occasion the expression of the intended product.

The expression plasmids may be assembled via three part ligation reactions as shown in FIG. 6. 15 ng (~0.05 pmole) of the assembled FIF gene (504 or 505 bp), 0.5 µg (~0.2 pmole) of the large Pst I-Bgl II fragment of pHKY10 and 0.2 µg (~0.3 pmole) of the appropriate promoter fragment were ligated and the mixture used to transform *E. coli* 294 (22). Plasmid DNA was prepared from individual transformants and analyzed by restriction mapping. Correct joining of the assembled gene to the promoter fragment should restore the Eco RI (lac) or the Xba I (trp) recognition sequences. The majority of the plasmids gave the expected restriction enzyme digestion patterns. Individual clones (12 containing the trp promoter and 12 containing the lac promoter) were grown and extracts prepared for interferon assay as described in Materials and Methods.

When assayed on human amnion (WISH) cells for antiviral activity by the CPE inhibition assay (1) five of the trp transformants were positive (each approximately equivalent); eleven of the lac transformants gave equivalent IF activities. Therefore, one transformant from each series (pFIFlac9 and pFIFtrp69) was selected for further study (Table 1). DNA sequence analysis demonstrated that the desired attachment of promoter to FIF structural gene had occurred in both cases. t,0210

The amounts of fibroblast interferon produced by pFIFlac9 and pFIFtrp69 are shown in Table 1. The trp promoter gave a FIF expression level measurably higher than did the lac promoter. In an attempt to further increase FIF expression levels, pFIFtrp69 was cleaved with Eco RI and two 300 base pair Eco RI fragments containing the trp promoter (30) were inserted. The resulting plasmid, pFIFtrp³⁶⁹, contains three successive trp promoters which read toward the FIF gene. The amount of FIF synthesized by *E. coli* K-12 strain 294/pFIF trp³⁶⁹ is 4-5 times that produced by pFIF trp 69 (Table 1). This is apparently due to the derepression of the trp promoter which occurs when trp repressor levels are titrated by the multiple copies of the trp operator.

The FIF produced by *E. coli* K-12 strain 294/pFIFtrp69 behaves like authentic human FIF. As shown in Table 2, its antiviral activity is about 30 times greater on human cells than bovine cells. In addition, the bacterially produced FIF is stable to treatment at pH 2 overnight and is not neutralized by rabbit antihuman leukocyte interferon antibodies (Table 3). t,0220 t,0221

Purification

The purification procedure for bacterial derived fibroblast is as follows:

1. Frozen cells are suspended in twelve times volume per weight with sucrose lysis buffer (100 mM Tris-HCl, 10 percent sucrose, 0.2M NaCl, 50 mM EDTA, 0.2 mM PMSF, pH 7.9) containing lysozyme at 1 mg ml⁻¹. The cell suspension is stirred for 1 hour at 4° C. and centrifuged. Fibroblast interferon activity remains in the supernatant.
2. Polyethyleneimine (5 percent v/v) is added to the sonicated supernatant to a final concentration of 0.5 percent (v/v). The solution is stirred for 1 hour at 4° C. and centrifuged. Interferon activity remains in the supernatant.
3. Solid ammonium sulfate is added to the polyethyleneimine supernatant to a final concentration of 50 percent saturation, stirred for 30 minutes at 4° C. and centrifuged. Interferon activity is in the 50 percent pellet.
4. The 50 percent ammonium sulfate pellet is suspended in one half the volume of the 50 percent ammonium sulfate suspension with Phosphate Buffered Saline (20 mM sodium phosphate 0.15M NaCl, pH 7.4). Polyethylene glycol 6000 (50 percent w/v in PBS) is added to a final concentration of 1½ percent (v/v), stirred at 4° C. for 2 hours and centrifuged. Interferon activity is in the pellet. The pellet is suspended in a minimal volume of sucrose lysis buffer and clarified by centrifugation.

This initial extraction procedure results in a purification of fibroblast interferon from 0.001 percent of the total protein to 0.05 percent of the total protein. This material can be further purified to homogeneity by the following column chromatography steps:

5. Affinity chromatography on Amicon Blue B in sucrose lysis buffer.
6. Anion exchange chromatography on QAE Sephadex in sucrose lysis buffer in the absence of 0.2M NaCl.
7. Size exclusion chromatography on Sephadex G-75 in sucrose lysis buffer.
8. Reverse phase high pressure liquid chromatography.

Parenteral Administration

FIF may be parenterally administered to subjects requiring antitumor or antiviral treatment. Dosage and dose rate may parallel that currently in use in clinical investigations of human derived materials, e.g., about (1-10)×10⁶ units daily, and in the case of materials of purity greater than 1 percent-

age, likely up to, e.g., 150×10^6 units daily. Dosages of bacterially obtained FIF could be significantly elevated for greater effect owing to the essential absence of human proteins other than FIF, which proteins in fibroblast-derived materials may act as pyrogens, exhibiting adverse effects, e.g., malaise, temperature elevation, etc.

As one example of an appropriate dosage form for essentially homogeneous bacterial FIF in parenteral form, 3 mg. FIF of specific activity of, say, 2×10^8 μ /mg may be dissolved in 25 ml. 5 percentage serum albumin (human) - USP, the solution passed through a bacteriological filter and the filtered solution aseptically subdivided into 100 vials, each containing 6×10^6 units pure interferon suitable for parenteral administration. The vials are preferably stored in the cold ($-20^\circ \text{C}.$) prior to use.

The compounds of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby the polypeptide hereof is combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation are described in Remington's *Pharmaceutical Sciences* by E. W. Martin, which is hereby incorporated by reference. Such compositions will contain an effective amount of the interferon protein hereof together with a suitable amount of vehicle in order to prepare pharmaceutically acceptable compositions suitable for effective administration to the host. One preferred mode of administration is parenteral.

BIBLIOGRAPHY

1. Stewart, W. E. II (1979) *The Interferon System*, Springer, N.Y.
 2. Knight, E. Jr. (1976) *Proc. Natl. Acad. Sci. U.S.A.* 73, 520-523.
 3. Berthold, W., Tan, C., and Tan, Y. H. (1978) *J. Biol. Chem.* 253, 5206-5212.
 4. Knight, E. Jr., Hunkapillar, M. W., Korant, B. D., Hardy, R. W. F. and Hood, L. E. (1980) *Science* 207, 525-526.
 5. Houghton, M., Stewart, A. G., Doel, S. M., Emtage, J. S., Eaton, M. A. W., Smith, J. C., Patel, T. P., Lewis, H. M., Porter, A. G., Birch, J. R., Cartwright, T., and Carcy, N. H. (1980) *Nucleic Acids Res.* 8, 1913-1931.
 6. Taniguchi, T., Sakai, M., Fujii-Kuriyama, Y., Muramatsu, M., Kobayashi, S., and Sudo, T. (1979) *Proc. Jan Acad.* B55, 464-469.
 7. Derynck, R., Content, J., DeClercq, E., Volckaert, G., Tavernier, J., Devos, R., and Fiers, W. (1980) *Nature* 285, 542-547.
 8. Clewell, D. B. (1972) *J. Bacteriol.* 110, 667-676.
 9. Maxam, A. M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
 10. Taylor, J. M., Ilmensee, R. and Summer, S. (1976) *Biochim. Biophys. Acta* 442, 324-330.
 11. Grunstein, M. and Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72, 3961-3965.
 12. Crea, R., Kraszewski, A., Hirose, T. and Itakura, K. (1978) *Proc. Natl. Acad. Sci. U.S.A.* 75, 5765-5769.
 13. Hirose, T., Crea, R. and Itakura, K. (1978) *Tetrahedron Letters* 28, 2449-2452.
 14. Crea, R. and Horn, T. (1980) *Nucleic Acids Res.* 8, 2331-2348.
 15. Pestka, S., McInnes, J., Havell, E. A. and Vileek, J. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72, 3898-3901.
 16. Green, M., Zehavi-Willner, T., Graves, P. N., McInnes, J. and Pestka, S. (1975) *Arch. Biochem. Biophys.* 172, 74-89.
 17. Gurdon, J. B., Lane, C. D., Woodland, H. R. and Marbaix, G. (1971) *Nature* 233, 177-182.
 18. Cavaliere, R. L., Havell, E. A., Vileek, J. and Pestka, S. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 3287-3291.
 19. Scheller, R. H., Dickerson, R. E., Boyer, H. W., Riggs, A. D. and Itakura, K. (1977) *Science* 196, 177-180.
 20. Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schimke, R. T. and Cohen, S. N. (1978) *Nature* 275, 617-624.
 21. Goeddel, D. V., Heyneker, H. L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D. G., Ross, M. J., Miozzari, G., Crea, R. and Seeburg, P. H. (1979) *Nature* 281, 544-548.
 22. Backman, K., Ptashne, M. and Gilbert, W., (1976) *Proc. Natl. Acad. Sci. U.S.A.* 73, 4174-4178.
 23. Hershfild, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. and Helinski, D. R. (1974) *Proc. Natl. Acad. Sci. U.S.A.* 71, 3455-3459.
 24. Galau, G. A., Britten, R. J. and Davidson, E. H. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 1020-1023.
 25. Birnboim, H. C. and Doly, J. (1979) *Nucleic Acid Res.* 7, 1513-1523.
 26. Kafatos, F. C., Jones, C. W. and Efstratiadis, A. (1979) *Nucleic Acids Res.* 7, 1541-1552.
 27. Denhardt, D. T. (1966) *Biochem. Biophys. Res. Comm.* 23, 641.
 28. Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A. D. (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76, 106-110.
 29. Miller, J. H. (1972) *Experiments in Molecular Genetics*, pp. 431-433, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 30. U.S. patent application of David V. Goeddel and Sidney Pestka U.S. Ser. No. 184909, filed Sep. 8, 1980 for Microbial Production of Mature Human Leukocyte Interferon, Attorney docket 1980/104, assigns Genentech, Inc. and Hoffmann-LaRoche, Inc.
 31. Taniguchi, T., Ohno, S., Fujii-Kuriyama, Y. and Muramatsu, M. (1980) *Gene* 10, 11-15.
 32. U.S. patent application Ser. No. 06/133,296 filed Mar. 24, 1980 by Dennis G. Kleid et al., assigned to Genentech, Inc.
 33. Klenow, H. and Henningsen, I. (1970) *Proc. Natl. Acad. Sci. U.S.A.* 65, 168-171.
- We claim:
1. A composition comprising water and a nonglycosylated polypeptide having the amino acid sequence of a mature human fibroblast interferon, said nonglycosylated polypeptide having a total of 165 or 166 amino acids and said composition being free of any glycosylated human fibroblast interferon.
 2. The composition of claim 1, said nonglycosylated polypeptide having the amino acid sequence
X-Ser-Tyr-Asn-Leu-Lcu-Gly-Phe-Lcu-Gln-Arg-Ser-Ser-Asn-Phe-Gln-Cys-Gln-Lys-Leu-Leu-Trp-Gln-Leu-Asn-Gly-Arg-Leu-Glu-Tyr-Cys-Leu-Lys-Asp-Arg-Met-Asn-Phe-Asp-Ile-Pro-Glu-Glu-Ile-Lys-Gln-Leu-Gln-Gln-Phe-Gln-Lys-Glu-Asp-Ala-Ala-Lcu-Thr-Ile-Tyr-Glu-Met-Leu-Gln-Asn-Ile-Phe-Ala-Ile-Phe-Arg-Gln-Asp-Ser-Ser-Ser-Tyr-Gly-Trp-Asn-Glu-Thr-Ile-Val-Glu-Asn-Leu-Leu-Ala-Asn-Val-Tyr-His-Gln-Ile-Asn-His-Leu-Lys-Thr-Val-Leu-Glu-Glu-Lys-Leu-Glu-Lys-Glu-Asp-Phe-Thr-Arg-Gly-Lys-Leu-Met-Ser-Ser-Leu-His-Leu-Lys-Arg-Tyr-Tyr-Gly-Arg-Ile-Lcu-His-Tyr-Leu-Lys-Ala-Lys-Glu-Tyr-Ser-His-Cys-Ala-Trp-Thr-Ile-Val-Arg-Val-Glu-Ile-Lcu-Arg-Asn-Phe-Tyr-Phe-Ile-Asn-Arg-Leu-Thr-Gly-Tyr-Leu-Arg-Asn,
wherein X is H or Met.
 3. The composition of claim 2, said nonglycosylated polypeptide having a formula molecular weight of about 20,027.

17

4. The composition of claim 1, 2 or 3, said composition being free of human proteins.

5. The composition of claim 1, 2 or 3, said composition containing a therapeutically effective amount of said nonglycosylated polypeptide and being suitable for parenteral administration. 5

18

6. The composition of claim 4, said composition containing a therapeutically effective amount of said nonglycosylated polypeptide and being suitable for parenteral administration.

* * * * *

**UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION**

PATENT NO. : 5,460,811

Page 1 of 6

DATED : October 24, 1995

INVENTOR(S) : David V. Goeddel and Roberto Crea

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

In column 5, line 663, please delete ($\alpha^{32}p$)dCTP and insert ($\alpha^{32}P$)dCTP.

In column 7, line 16, please delete 500 and insert 550.

In column 10, line 61, please delete GAATCATG and insert GAATTCATG.

In column 11, line 52, please delete t,0170 and insert

5' CTAGA—
3' T—

5' CTAGA—
3' TCT—

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

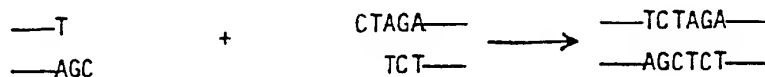
PATENT NO. : 5,460,811

DATED : October 24, 1995

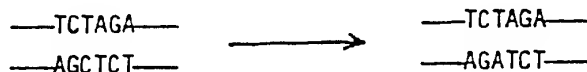
Page 2 of 6

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 11, line 67, please delete t,0180 and insert



In column 12, line 6, please delete t,0181 and insert



UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

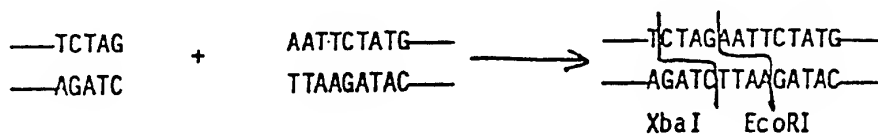
Page 3 of 6

INVENTOR(S) : David V. Goeddel and Roberto Crea
In column 12, line 51, please delete t,0190 and insert

XbaI filled in

EcoRI filled in

HGH gene initiation -



UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

Page 4 of 6

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 13, line 67, please delete t,0210 and insert

Table 1. Interferon activity in extracts of E. coli

| <u>E. coli</u> K-12 strain 294 transformed by: | Cell density (cells/ml) | IF Activity (units/l culture) | FIF molecules per cell |
|--|----------------------------|----------------------------------|---------------------------|
| pBR322 | 3.5×10^8 | - | - |
| pFIFlac9 | 3.5×10^8 | 9.0×10^6 | 2,250 |
| pFIFtrp69 | 3.5×10^8 | 1.8×10^7 | 4,500 |
| pFIFtrp ³⁶⁹ | 3.5×10^8 | 8.1×10^7 | 20,200 |

**UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION**

PATENT NO. : 5,460,811

DATED : October 24, 1995

Page 5 of 6

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 14, line 21, please delete t,0220 and insert

Table 2. Interferon activities measured on different cell types

| <u>Cells</u> | <u>Interferon Activity (units/ml)</u> | | |
|---------------|---------------------------------------|--------|--|
| | LeIF | FIF | E. coli K-12 strain 294/pFIFtrp69 extract |
| Human amnion | 20,000 | 10,000 | 1280 |
| Bovine kidney | 13,000 | 400 | 40 |

LeIF and FIF were NIH standard solutions having 20,000 units/ml and 10,000 units/ml respectively. Assays were performed as described in Materials and Methods.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,460,811

DATED : October 24, 1995

Page 6 of 6

INVENTOR(S) : David V. Goeddel and Roberto Crea

In column 14, line 21, please delete t,0221 and insert

Table 3. Comparison of activities of extracts from E. coli K-12 strain 294/pFIFtrp69 with standard human leukocyte and fibroblast interferons

| | <u>Interferon Activity (units/ml)</u> | | |
|----------------------------------|---------------------------------------|------------|--|
| | <u>LeIF</u> | <u>FIF</u> | <u>E. coli K-12 strain 294/pFIFtrp69</u> |
| untreated | 1000 | 1000 | 1000 |
| pH2 | 1000 | 1000 | 1000 |
| rabbit antihuman LeIF antibodies | <16 | 1000 | 1000 |

Experimental procedures described in Materials and methods. Assayed by CPE inhibition using WISH cells/Sindbis virus.

Signed and Sealed this
Ninth Day of April, 1996

Attest:



BRUCE LEHMAN

Attesting Officer

Commissioner of Patents and Trademarks

TAB VV

THIS WEEK IN SCIENCE

of specific chromosomal regions, including gene clusters encompassing neuron-specific genes, some of which do not themselves contain REST/NRSF response elements. Σ

Step-by-Step Assembly

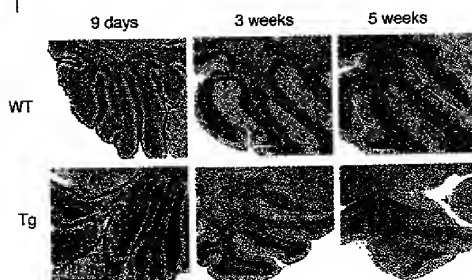
A family of RNA-protein complexes, known as the uridine-rich small nuclear ribonucleoproteins (U snRNPs), form the core of the spliceosome that excises introns and ligates exons to form messenger RNA. A functional deficiency in the survival of motor neurons (SMN) protein results in spinal muscular atrophy, a disease in which motor neurons of the spinal cord degenerate. Pellizzoni *et al.* (p. 1775) show that the SMN complex (of which the SMN protein is a part) serves to assemble in an orderly fashion the protein ingredients of U snRNPs (the Sm proteins) onto the U snRNAs. It first binds the Sm proteins, then the U snRNAs, and finally puts them together in an adenosine-triphosphate (ATP-dependent) reaction.

Neurotrypsin and Mental Retardation

Inherited mental retardation (MR) is often linked to abnormalities on the X chromosome or to abnormalities in brain development or other clinically identifiable features, but in most cases none of these attributes are present. An analysis of such nonsyndromic MR patients by Molinari *et al.* (p. 1779) revealed an association with the mutation of the serine protease neurotrypsin. In situ hybridization studies of the expression of neurotrypsin during normal development revealed that it is expressed in parts of the brain associated with learning and memory and first appears at 44 days of development. Immunoelectron microscopy localized neurotrypsin at presynaptic nerve endings. Although this mutation does not appear to be a common cause of MR, further studies may yield insights into the pathways leading to these diseases.

Neurotoxic, Cytosolic Prion Proteins

The characteristics of the prion protein, which has been linked to a variety of neurodegenerative disorders, is the subject of two reports. Ma and Lindquist (p. 1785) reveal how inhibition of the proteasome machinery in cells that produce prion proteins can lead to the accumulation of prion isoforms in the cytosol, and how, under certain conditions, misfolded self-perpetuating isoforms can be generated de novo. Ma *et al.* (p. 1781) examined the effects of retrograde-transported or cytosolically expressed prion protein in a transgenic mouse model and in neuronal cell lines. Cytosolic prion protein was highly neurotoxic, and mice engineered to possess cytosolic prion protein developed severe ataxia, cerebellar degeneration, and gliosis. Σ



The Hydrogen Economy of the Gut

A major cause of gastritis, peptic ulcers, and certain cancers is the bacterium *Helicobacter pylori*. Olson and Maier (p. 1788) show that the colonization success of this common pathogen is boosted by hydrogen gas produced by other intestinal occupants. Molecular hydrogen in the mucous lining of the stomach stimulates the pathogen to produce more of a constitutive enzyme, hydrogenase, required to harvest the energy through a series of heme-containing electron carriers.

Glycosylation in Bacteria

N-linked glycosylation is a common posttranslational modification of membrane and secretory proteins in eukaryotes. However, the only bacterium known to modify proteins in this way is *Campylobacter jejuni*, and thus many eukaryotic proteins generated in bacterial systems are of limited use because they lack appropriate modifications. Wacker *et al.* (p. 1790) have transplanted the N-linked glycosylation machinery from *C. jejuni* to *Escherichia coli*, which should enhance the opportunities for the large-scale production of appropriately modified proteins in bioreactors.

NanoDrop UV/Vis Spectrophotometer

You'll try it
for its 1 μ l
capability.

But it's so easy, you'll
use it for everything.

No cuvettes or capillaries

At-strength samples

without dilution

Very small footprint

The NanoDrop™ ND-1000 spectrophotometer is so quick and easy, you'll use it for all your samples. Simply pipette a sample directly onto the measurement surface. The sample is drawn into a column between two optical fibers to establish the measurement path. After measurement is complete, a simple wipe and you're ready for your next sample.

NanoDrop ND-1000 for all UV/Vis measurements, including:

- nucleic acids (even at 3500 ng/ μ l)
- fluorescent-dye labeled microarray samples
- proteins
- cell cultures
- full spectrum analysis (220-750 nm)

Visit our website for demonstration video, testimonials, and full specifications.

www.nanodrop.com

NanoDrop™ Technologies

Rockland, Delaware USA
Telephone (302) 984-0334

©2002 NanoDrop Technologies

TAB WW

bonuclease, J.

VIRIOLOGY 63, 475-483 (1975)

RNAGEL, M. J.
ed from beans
1400.AKSENA, K. N.
d properties of
stunt virus.

raw-Hill, New

Suppression of Human Interferon Production by Inhibitors of Glycosylation

EDWARD A. HAVELL, JAN VILČEK, ERNESTO FALCOFF¹, AND BRIAN BERMAN

Department of Microbiology, New York University School of Medicine, New York, New York 10016

Accepted September 27, 1974

Treatment of diploid human foreskin cell cultures with either 2-deoxy-D-glucose or D-glucosamine resulted in dose-dependent inhibition of interferon production induced with polyinosinate-polycytidylate [poly(I)·poly(C)]. This inhibitory action was readily demonstrable even if the addition of the inhibitor was delayed until 1.5 hr after induction. The appearance of intracellular interferon activity was also inhibited, suggesting that synthesis of biologically active interferon, and not its secretion from the cell, was the main target of the inhibitory action. The inhibitory action of 2-deoxy-D-glucose on interferon production was completely reversible by D-mannose while the action of D-glucosamine was only slightly affected by the addition of mannose. Interferon produced in the presence of a partially inhibitory concentration of 2-deoxy-D-glucose was more heat-labile than control interferon. Residual active interferon made in the presence of D-glucosamine was less efficiently neutralized by anti-interferon antibody than control interferon. The results are compatible with the idea that the two inhibitors suppress the production of biologically active interferon by virtue of their interfering with the proper glycosylation of the molecule.

INTRODUCTION

The evidence that interferons are glycoproteins has recently been reviewed by Weil and Dorner (1973). Schonke *et al.* (1970) were the first to show that treatment of rabbit interferon preparations with neuraminidase, while not affecting biological activity, abolished the characteristic charge heterogeneity of interferon demonstrable by isoelectric focusing. This finding led the authors to propose that rabbit interferon is a glycoprotein with a variable content of terminal sialic acid residues.

Dorner *et al.* (1973) carried this work a step further by showing that rabbit interferon treated with neuraminidase (which banded in a single peak at pH 6.3) regained its original charge heterogeneity after reincorporation of sialic acid residues by the use of sialyl transferase from rat liver. These authors also showed that the carbo-

hydrate portion of rabbit interferon contains at least one terminal sequence, sialic acid-galactose-.

The multiplication of several enveloped viruses was shown to be inhibited by 2-deoxy-D-glucose or D-glucosamine. The list of viruses shown to be inhibited by one or both of these compounds includes influenza (Kilbourne, 1959; Gandhi *et al.*, 1972; Klenk *et al.*, 1972), Sindbis and Semliki Forest (Kaluza *et al.*, 1972, 1973), Newcastle disease (Samson and Fox, 1974; Scholtissek *et al.*, 1974), herpes simplex (Courtney *et al.*, 1973) and avian sarcoma (Hunter *et al.*, 1974) viruses. Most available evidence suggests that this suppression of virus multiplication is due to selective interference with normal glycosylation of viral envelope glycoproteins resulting in the derangement of their processing.

These reports prompted us to examine the effect of 2-deoxy-D-glucose and D-glucosamine on interferon synthesis, to determine if, in this instance too, these

¹ Permanent address: Institut du Radium, Paris, France. Recipient of Eleanor Roosevelt Fellowship from International Union Against Cancer.

compounds might exert a selective effect on glycosylation, and to obtain new information about the function of the carbohydrate portion of the interferon molecule.

MATERIALS AND METHODS

Cell cultures. A human diploid cell strain (FS-4), derived in this laboratory from the foreskin of a single neonate and identified as a high interferon producer (Vilček and Havell, 1973), was used throughout. All experiments were carried out in cultures grown in 60-mm plastic petri dishes, unless specified otherwise. Each dish was seeded with a total of 200,000 cells in 5 ml of medium consisting of Eagle's minimal essential medium (MEM) buffered with *N*-2-hydroxyethyl-piperazine - *N'* - 2 - ethanesulfonic acid (hepes, 6 mM), *N*-tris-hydroxymethyl-methyl glycine (tricine, 13 mM) and sodium bicarbonate (1.1 g/liter), supplemented with 5% heated fetal calf serum. The medium was changed for fresh growth medium on the 6th day and the cultures were used for experiments 11-13 days after seeding. The incubation was in a humidified CO₂ incubator at 36°, unless specified otherwise.

Chemicals. Polyinosinate-polycytidyate [poly(I)·poly(C)] was supplied by the Antiviral Substances Program, Infectious Disease Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD. D-Glucosamine, 2-deoxy-D-glucose, D-mannose, and *N*-acetyl-D-glucosamine were purchased from Sigma Chemical Co., St. Louis, MO. Solutions of the sugars were prepared in water, sterilized by membrane filtration, and stored at 4°. Cycloheximide was obtained from the Upjohn Co., Kalamazoo, MI, and actinomycin D from Calbiochem, Los Angeles, CA.

Interferon titrations. The microassay for human interferon was described in detail earlier (Havell and Vilček, 1972). In short, wells of the Micro Test II plastic tissue culture plates (Falcon Plastics, Los Angeles, CA, or Linbro, New Haven, CT) were filled with 100 μ l of MEM containing 5% fetal calf serum and duplicate serial two-fold dilutions of the assayed materials were

prepared with the aid of an Eppendorf micropipette. To each well was then added a suspension of human foreskin cells (30,000 cells/well in 100 μ l of medium, as above). After 18-20 hr of incubation, 1,000 PFU of the Indiana type vesicular stomatitis virus was added to each well and a final reading of the test was taken by microscopic examination about 48 hr after virus inoculation. The highest dilution of the assayed sample which protected at least 50% of the cell sheet from the cytopathic action of the virus was taken as the end point. Each assay included the titration of an interferon standard that had been calibrated against the 69/19 reference standard for human interferon (obtained from Antiviral Substances Program, Infectious Disease Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD). All interferon titers throughout the paper are expressed in terms of reference units/ml.

Preparation of high-titered interferon by superinduction. Dish-cultures of FS-4 cells were incubated with poly(I)·poly(C) (100 μ g/ml) and cycloheximide (50 μ g/ml) in 2 ml of serum-free MEM. Actinomycin D (final concentration 1 μ g/ml) was added to the cultures 4 hr after the onset of this treatment. At 6 hr (i.e., 2 hr after the addition of actinomycin), medium with the inducer and the inhibitors was removed, the cells were thoroughly washed and replenished with 2 ml of MEM containing 2% fetal calf serum. Culture fluids were harvested 30 hr after induction. The incubation temperature throughout this procedure was 34° (Vilček and Havell, 1973).

Preparation of rabbit antiserum against interferon. Interferon prepared in serum-free MEM by superinduction of roller-bottle cultures of FS-4 cells and concentrated by ultrafiltration was employed for immunization. Three New Zealand white female rabbits received a total of three subcutaneous injections in Freund's complete adjuvant, seven injections in Freund's incomplete adjuvant and, finally, eight injections without adjuvant. A total of 32,000 units of interferon was administered with each injection. Serum was ob-

obtained from blood drawn from the ear veins, heated at 56° for 30 min, and absorbed for 1 hr with normal FS-4 cells (about 4×10^6 cells/ml serum) in order to remove antibody against cellular surface components.

Neutralization assay of interferon antiserum. An adaptation of the microassay for human interferon was employed for the titration of neutralizing activity of anti-interferon sera. Micro Test II plastic tissue culture plates were first seeded with a suspension of human foreskin cells (25,000 cells/well) in 0.2 ml of MEM containing 5% fetal calf serum. After 18 hr incubation the medium was aspirated from the wells and replaced with 0.2 ml of a mixture of serial twofold dilutions of the antiserum plus a constant dose of the tested interferon sample (10 units/ml, final concentration). Prior to adding it to the wells, the mixture of antiserum dilutions and interferon had been incubated for 1 hr at 37° in order to allow neutralization to take place. After 18–20-hr incubation the antibody-interferon mixture-containing media were aspirated and replaced with 0.25 ml of MEM containing 5% fetal calf serum and 1,000 PFU of vesicular stomatitis virus. The assay was read 48 hr after virus inoculation. The neutralizing titer was the highest dilution of the serum that inhibited the antiviral action of interferon as determined by the presence of cytopathic effect. Each sample was assayed in duplicate or triplicate, with adequate controls for antiserum, interferon and virus.

Incorporation of [14 C]leucine. FS-4 cultures in 35-mm plastic petri dishes were incubated with 0.4 μ Ci of uniformly labeled L-[14 C]leucine (0.25 mCi/0.12 mg; New England Nuclear Corp., Boston, MA) in 1 ml complete MEM with 2% fetal calf serum. Incorporation was stopped after 60 min by washing the cells three times with ice-cold buffered saline. The cells were then lysed in 2 ml of 0.1 N NaOH and heated for 30 min at 36°. Lysates were made to 10% with trichloroacetic acid (TCA); the resulting precipitates were collected on Whatman GF/C glass-fiber filters (Reeve Angle, Clifton, NJ), washed four

times with 5 ml each of 5% TCA and rinsed once with 95% EtOH. Acid-insoluble radioactivity was measured in a liquid scintillation counter and expressed as cpm/culture. Preliminary experiments showed that expressing the results as cpm/mg of protein did not improve the accuracy of measurement.

Treatments with 2-deoxy-D-glucose or D-glucosamine. These were always carried out in complete MEM, i.e., in the presence of 5.5 mM D-glucose. Glucose was retained in the medium because it is the major carbon and energy source.

RESULTS

Dose Response and Kinetics of Inhibition

Treatment of FS-4 cell cultures with either 2-deoxy-D-glucose or D-glucosamine resulted in a dose-dependent inhibition of the interferon yield, as determined in culture fluids collected 6 hr after induction with poly(I)·poly(C). Both compounds were about equally effective in inhibiting the production of biologically active interferon (Fig. 1).

Previous studies showed that interferon production in this system is sensitive to inhibition by actinomycin D only within the first hour after induction, indicating that the bulk of interferon mRNA synthe-

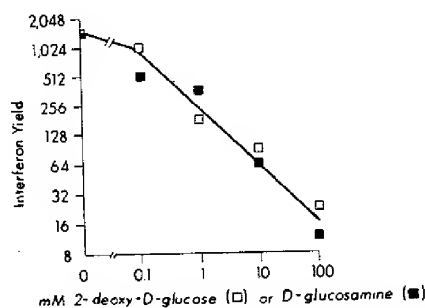


FIG. 1. Effect of various concentrations of 2-deoxy-D-glucose and D-glucosamine on the interferon yield. Cultures of FS-4 cells were stimulated to produce interferon by 30-min incubation with poly(I)·poly(C) (100 μ g/ml) in 2 ml of MEM. The cultures were thoroughly washed thereafter and replenished with 2 ml of MEM plus 2% fetal calf serum containing the inhibitors as indicated. All culture fluids were collected 6 hr after exposure to the inducer, dialyzed and assayed for interferon.

sis is completed by 1.5 hr (Vilček and Havell, 1973, 1974). When the time of addition of 2-deoxy-D-glucose to cultures was delayed until 1.5 hr after exposure of cells to poly(I)·poly(C) the resulting inhibition of interferon production was similar to that seen in cultures that had been treated with the inhibitor from the time of exposure to the inducer (Table 1). Reversal of the inhibition at 8 hr after induction did not result in a subsequent release of appreciable quantities of interferon. These results indicate that 2-deoxy-D-glucose interferes irreversibly with a posttranscriptional event in interferon synthesis.

Inhibition of Intracellular Appearance of Interferon

In cells stimulated with poly(I)·poly(C), intracellular interferon is detectable about 30 min prior to its appearance in the extracellular fluid (Tan *et al.*, 1971; Ng *et al.*, 1972). The process of interferon release is temperature dependent and energy dependent but, unlike interferon synthesis, it does not require protein synthesis (Tan *et al.*, 1972).

Treatment with 2-deoxy-D-glucose reduced the amount of interferon recoverable from extracts of induced cells (Fig. 2).

TABLE 1

EFFECT OF VARYING THE TIME OF TREATMENT WITH 2-DEOXY-D-GLUCOSE ON THE INTERFERON YIELD

| Time of treatment (hr) ^a | Interferon yield at hr ^b | |
|-------------------------------------|-------------------------------------|----|
| | 8 | 24 |
| None | 768 | 96 |
| 0-8 | 64 | 16 |
| 1.5-8 | 96 | 12 |

^a All cultures were exposed at 0 hr to poly(I)·poly(C) (100 µg/ml) in 2 ml of MEM and incubated for 90 min. Thereafter, all cultures were thoroughly washed and replenished with 2 ml of MEM containing 2% fetal calf serum. In addition, cultures were treated with 2-deoxy-D-glucose (6 mM) for the times indicated.

^b Culture fluids were collected 8 hr after induction, the cultures were washed, replenished with fresh maintenance medium (2 ml of MEM with 2% fetal calf serum) and incubated until 24 hr. The harvested fluids were dialyzed prior to assaying for interferon.

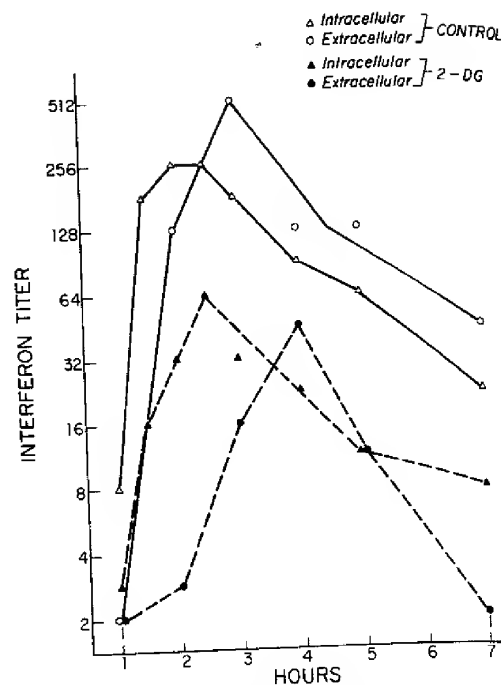


FIG. 2. Effect of 2-deoxy-D-glucose (2-DG) on the intracellular appearance of interferon and its secretion. Cultures were induced with poly(I)·poly(C) for 30 min as described in Fig. 1. One group of cultures was then replenished with 2 ml of medium containing 2-DG (10 mM) while the other group received maintenance medium. For the determination of intracellular interferon levels, media were removed from the cultures at the indicated intervals, the cell sheet was washed twice with ice-cold saline and subjected to five cycles of rapid freezing (using a dry ice-alcohol bath) and thawing at 37° in 2 ml of MEM with 2% fetal calf serum. The homogenates were dialyzed prior to interferon assay. For the determination of extracellular interferon levels, medium from each of the same two treated and control plates was collected at 1, 2, 3, 4, 5 and 7 hr. After collecting the medium, the cells were quickly washed once with warm (37°) medium, replenished with 2 ml of fresh warm maintenance medium (with or without 2-DG added) and immediately returned to the incubator until the next interval. All collected fluids were dialyzed prior to interferon assay. Since all preceding culture fluid samples had been collected at hourly intervals, the 7-hr extracellular interferon titers were divided by 2 to correct for the longer interval that had elapsed before the last harvesting period. Thus all extracellular interferon titers actually represent the cumulative yields/hr, whereas intracellular interferon titers represent the net amount of interferon recovered at the particular interval from the cell extracts, after dilution of the extracts in the 2 ml of medium added before freeze-thawing.

ar } CONTROL
lar }
lar } 2-DG



Thus the inhibitory action of this compound could not be attributed simply to an interference with the release process. However, in addition to the reduction in the titer of intracellular interferon, treatment with 2-deoxy-D-glucose also caused a slight delay in the extracellular appearance of the interferon that escaped intracellular inhibition, suggesting some disturbance of the release mechanism. Note that at 2 hr the ratio intracellular/extracellular interferon is 2 in the control cultures compared to 10 in the treated group.

Treatment with D-glucosamine also inhibited the appearance of both intracellular and extracellular interferon, whereas N-acetylglucosamine exerted no significant effect on either intracellular or extracellular interferon titers (Table 2).

Reversal of the Inhibitory Effect of 2-Deoxy-D-Glucose by Mannose

Kaluza *et al.* (1973) showed that the addition of D-mannose specifically interfered with the inhibitory action of 2-deoxy-D-glucose on the synthesis of infectious Semliki Forest virus. Earlier, Kaluza *et al.* (1972) demonstrated that a number of sugars, including mannose, did not pro-

duce a reversal of the inhibitory effect of D-glucosamine on the multiplication of enveloped RNA viruses.

At concentrations of 5 mM, 2-deoxy-D-glucose and D-glucosamine inhibited [¹⁴C]leucine incorporation by 60 and 46%, respectively (Fig. 3). [In other experiments, a similar degree of inhibition of protein synthesis by these compounds was seen in cells that had also been induced with poly(I)·poly(C).] At the same concentrations both of these compounds inhibited the interferon yield by 90%. The addition of mannose afforded a partial reversal of the inhibition of protein synthesis by 2-deoxy-D-glucose, while not affecting [¹⁴C]leucine incorporation in the presence of glucosamine. Even at the lowest concentration employed, mannose caused a complete reversal of the inhibitory effect of 2-deoxy-D-glucose on interferon production, while exerting a much less pronounced reversal of the action of glucosamine. Increasing the concentration of glucose in the medium caused only partial reversal of the inhibitory effect of 2-deoxy-D-glucose on interferon production (data not shown here).

These results are in general agreement with those of Kaluza *et al.* (1973) who showed that, although the cellular uptake of 2-deoxy-D-glucose was more strongly inhibited by excess glucose than by mannose, mannose had a more pronounced inhibitory effect on the incorporation of labeled 2-deoxy-D-glucose into acid-insoluble material. They proposed that 2-deoxy-D-glucose exerts its inhibitory effect on the multiplication of enveloped viruses mainly by being incorporated into viral glycoproteins instead of mannose.

The results of this experiment also show the disparity between the effect of these compounds on [¹⁴C]leucine incorporation and interferon production. In the absence of mannose, amino acid incorporation was inhibited less markedly than interferon production. However, in the presence of various concentrations of mannose, 2-deoxy-D-glucose still afforded a 33-23% inhibition of cellular protein synthesis while interferon production was increased

TABLE 2
EFFECTS OF D-GLUCOSAMINE OR N-ACETYLGLUCOSAMINE ON THE INTRACELLULAR AND EXTRACELLULAR APPEARANCE OF INTERFERON

| Treatment ^a | Interferon yield | | | |
|------------------------|-----------------------------|----------------------------------|-----|-----|
| | Intra-cellular ^b | Extracellular at hr ^c | | |
| | | 3 | 4 | 6 |
| None | 64 | 1,536 | 384 | 128 |
| D-glucosamine | 12 | 128 | 64 | 64 |
| N-acetylglucosamine | 64 | 1,024 | 384 | 96 |

^a Cultures were incubated with poly(I)·poly(C) for 30 min as described in Fig. 1. Maintenance medium contained D-glucosamine (46 mM), N-acetylglucosamine (45 mM), or no addition throughout the rest of the experiment.

^b Determined in cell extracts prepared at 3 hr after induction, as described in Fig. 2.

^c See Fig. 2. All results are actual interferon titers of the samples not corrected to yields/hr.

(2-DG) on the and its secre- (I)·poly(C) for up of cultures ium containing ceived mainte- of intracellular l from the cul- cell sheet was d subjected to dry ice-alcohol MEM with 2% e dialyzed prior tion of extracel- ach of the same lected at 1, 2, 3, edium, the cells (37°) medium, m maintenance d) and immedi- he next interval. or to interferon aid samples had : 7-hr extracellu- to correct for the before the last llular interferon lative yields/hr. rs represent the at the particular d dilution of the ed before freeze-

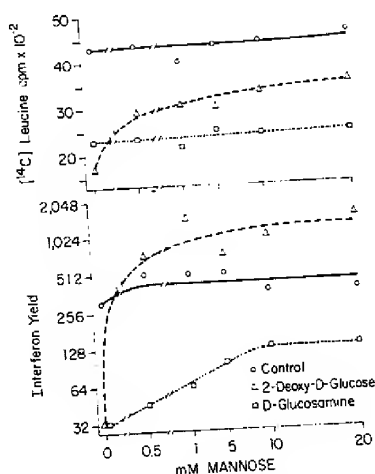


FIG. 3. Comparison of the effects of 2-deoxy-D-glucose and D-glucosamine on $[^{14}\text{C}]$ leucine incorporation and interferon production; reversal by mannose. $[^{14}\text{C}]$ leucine incorporation: Cultures were first incubated for 2 hr with 2-deoxy-D-glucose (5 mM), glucosamine (5 mM) or no inhibitor plus the concentration of mannose indicated. The culture medium consisted of complete MEM with 2% fetal calf serum. $[^{14}\text{C}]$ leucine (0.4 $\mu\text{Ci}/\text{culture}$) was added to the cultures at 2 hr after the onset of treatment. The labeling was allowed to proceed in the presence of the inhibitors and/or mannose for 1 hr. Acid-insoluble radioactivity, expressed as cpm/culture, was determined as described in Materials and Methods. Interferon yield: Stimulation of cultures with poly(I)·poly(C) was for 30 min, as described in Fig. 1. The cultures were then incubated with 2-deoxy-D-glucose (5 mM), glucosamine (5 mM) or no inhibitor plus the indicated concentration of mannose. Culture fluids were harvested at 24 hr, dialyzed and assayed for interferon.

even above control levels. It was demonstrated earlier than partial inhibition of protein synthesis may augment interferon yields, presumably through preferential inhibition of synthesis of a posttranscriptional repressor that exerts negative control over interferon production (Vilček, 1970).

Altered Characteristics of Interferon Made in the Presence of Inhibitors

The carbohydrate moiety is thought to be responsible for the characteristic thermal stability of many glycoproteins (Spiro, 1970, 1973). If treatment with 2-deoxy-D-glucose or D-glucosamine resulted in decreased or altered glycosylation, then one might expect that even the active inter-

feron made in the presence of a partially inhibitory concentration of one of these drugs would show altered thermal stability. This consideration rests on the assumption that a minor change in carbohydrate content would tend to affect thermal stability rather than biological activity.

To test this idea, a group of cultures was induced with poly(I)·poly(C) and one-half of the cultures was treated with 2-deoxy-D-glucose. In order to increase the yield of interferon, both sets of cultures were "superinduced" by sequential treatment with cycloheximide and actinomycin D as described (Vilček and Havell, 1973). Although the yield of interferon from the 2-deoxy-D-glucose-treated cultures was only 17% of the yield from control cultures, the activity (7,168 units/ml) was sufficiently high for the analysis of thermal inactivation kinetics at 56°. Since the rate of thermal inactivation of interferons is known to be affected by changes in hydrogen ion concentration (Marshall *et al.*, 1972), both preparations were first dialyzed against phosphate-buffered saline, pH 7.4.

A comparison of the two curves (Fig. 4) shows that inactivation of the interferon made in the presence of 2-deoxy-D-glucose proceeded at a somewhat faster rate. The reproducibility of this finding was confirmed in three separate experiments, employing two different batches of interferons from control and 2-deoxy-D-glucose-treated cultures.

The availability of interferon antiserum made it possible to compare the neutralization of control interferon and interferon made in the presence of the inhibitors. The neutralizing titer against interferon produced in the presence of partially inhibitory concentrations of D-glucosamine was consistently 4–8 times lower than against control interferon preparations produced in parallel either in untreated cultures or in cultures treated with *N*-acetylglucosamine (Table 3). It should be noted that *N*-acetylglucosamine showed no inhibitory effect on interferon production. The neutralizing titer of the same antiserum against interferon made in the presence of 2-deoxy-D-glucose was not significantly different from

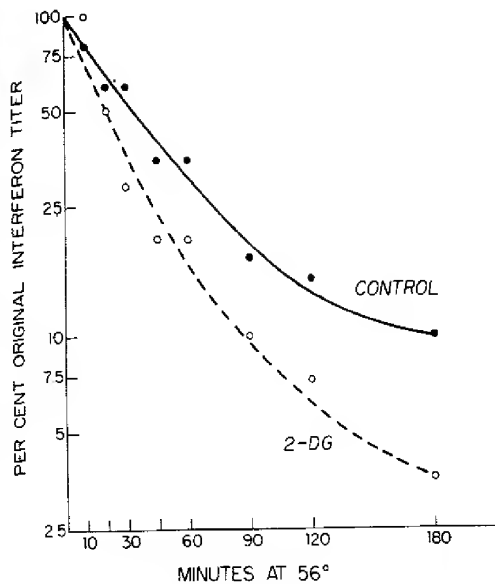


FIG. 4. Comparison of the rate of inactivation at 56° of interferon produced in the presence of 2-deoxy-D-glucose (2-DG) and control interferon. High-titered interferon was prepared in FS-4 cells superinduced by sequential treatment with cycloheximide and actinomycin D as described in Materials and Methods. One-half of the cultures was also incubated with 2-DG (6 mM) during the entire period of interferon production. Both the control interferon and 2-DG interferon were dialyzed against two changes of phosphate-buffered saline, pH 7.4. Aliquots of both preparations were heated in a 56°-water bath. Samples were taken at the indicated intervals, diluted in ice cold medium and assayed for interferon. The initial interferon titers/ml of the unheated preparations were 40,960 (control) and 7,168 (2-DG).

its titer against control interferon (data not shown here).

The decreased neutralizability of interferon from glucosamine-treated cultures could be the result of one of the following two events: either the biologically active interferon in these preparations is altered structurally and therefore has a lower affinity for antibody, or, alternatively, these interferon preparations contain an excess of interferon molecules that, while biologically inactive, are antigenic and thus capable of competing for antibody binding sites with biologically active interferon.

DISCUSSION

The inhibition of the production of biologically active interferon by 2-deoxy-D-

glucose and D-glucosamine could be the result of (1) interference with the energy supply of the cell, (2) some secondary effect due to depression in overall protein synthesis, or (3) inhibition of glycosylation or improper glycosylation of the interferon polypeptide.

Both D-glucosamine and 2-deoxy-D-glucose were shown to cause a decrease in the intracellular pool of several nucleotides (Letnansky, 1964; Bekesi and Winzler, 1969; Keppler *et al.*, 1970; Scholtissek, 1971). Although this action would be expected to result in a decrease of cellular energy supply, available results do not support the view that the effects of the antimetabolites can be explained simply by lack of energy. In chick cells treated with 2-deoxy-D-glucose, Kaluza *et al.* (1973) found no decrease in the energy charge of the adenylate pool. Our finding that [¹⁴C]leucine incorporation into total cellular protein was decreased by only about 50% is also not consistent with such an interpretation. Even prolonged treatment of cultures with the antimetabolites

TABLE 3
NEUTRALIZATION OF INTERFERON MADE IN THE PRESENCE OF D-GLUCOSAMINE BY ANTI-INTERFERON ANTIBODY

| Interferon preparation ^a | Neutralizing titer of antiserum ^b in Expt. no. | | |
|-------------------------------------|---|-----------------|-----|
| | 1 | 2 | 3 |
| Control | 560 | 535 | 819 |
| D-glucosamine | 66 | 120 | 205 |
| N-acetylglucosamine | 860 | ND ^c | ND |

^a Produced as described in Fig. 1, except that culture fluids were collected at 3 hr after induction with poly(I)·poly(C). The preparations used in Expt. no. 1 and 2 were produced in media that contained D-glucosamine (4.6 mM), N-acetylglucosamine (4.5 mM) or no addition and the respective interferon titers of the undiluted preparations were 128, 1,536 and 768 units/ml. The preparations used in Expt. no. 3 were the 3 hr extracellular fluids listed in Table 2. All samples were dialyzed against two changes of phosphate-buffered saline and one change of MEM before use in the neutralization assay.

^b Against 10 units of interferon/ml as described in Materials and Methods.

^c Not done.

was not toxic and the cells continued to divide, albeit at a reduced rate (unpublished data).

Although both compounds caused a detectable inhibition of labeled amino acid incorporation, this reduction in total cellular protein synthesis could hardly account for the inhibitory effect on interferon production. The strongest arguments against such interpretation come from the results of the experiment shown in Fig. 3, demonstrating that the degree of inhibition of overall protein synthesis is not directly related to the magnitude of the interferon response.

The specific nature of the inhibitory effect is also strongly supported by the finding that the addition of as little as 0.5 mM mannose completely reversed the inhibitory action on interferon production of a tenfold excess of 2-deoxy-D-glucose. This supports the view expressed earlier that 2-deoxy-D-glucose acts as an antimetabolite of mannose and that it is perhaps incorporated into glycoproteins instead of mannose (Kaluza *et al.*, 1973).

Finally, two direct lines of evidence suggest that treatment with both compounds may result in the synthesis of structurally altered interferon molecules: the decreased thermal stability of interferon made in the presence of 2-deoxy-D-glucose and the apparent lower affinity for antibody of interferon made in cells treated with D-glucosamine. Although direct proof is lacking, it is certainly conceivable that these altered characteristics are the result of either incomplete or improper glycosylation of the interferon molecule.

As with most other glycoproteins (Spiro, 1970, 1973), the functional role of the carbohydrate moiety of interferons remains obscure. Two hypotheses have attempted to account for the general significance of the carbohydrate portions of glycoproteins. Eylar (1966) proposed that proper glycosylation may be the chemical passport for export from the synthesizing cell. The idea that the carbohydrate moiety is a prerequisite for secretion was challenged by Winterburn and Phelps (1972) who pointed out that many secreted proteins (e.g., serum albumin) are not glycosylated. Instead, the

latter authors proposed that the sugar portion may be important in determining the extracellular fate and function of the molecule, such as the recognition of target cells. It should be pointed out that the two hypotheses are not mutually exclusive.

The inhibitory effect of 2-deoxy-D-glucose and glucosamine on interferon production can not be fully explained by selective inhibition of the secretion process alone, since intracellular levels of interferon were also suppressed. However, the results shown in Fig. 2 do not rule out the possibility that a lack of proper glycosylation may partially impede the release of interferon from producing cells.

It remains to be explained why proper glycosylation may be required for the synthesis of biologically active interferon. Although it seems attractive to speculate that the carbohydrate portion is important in the interaction of interferon with a cellular receptor site on target cells, this interpretation appears unlikely in view of two findings: it was shown that rabbit interferon retains full biological activity after removal of sialic acid (Schonne *et al.*, 1970; Dorner *et al.*, 1973), and enzymatic removal of about 50% of total carbohydrate from a preparation of human interferon also failed to reduce biological activity (C. B. Anfinsen, personal communication). Although these treatments resulted only in partial removal of carbohydrate, it seems likely that if indeed the carbohydrate portion determines the interaction with a cellular receptor site, the specificity of this reaction would depend on the exposed terminal carbohydrate residues. It is possible, however, that some parts of the core portion of the carbohydrate chains may be important for conformational stability of the interferon molecule which, in turn, might be required for full biological activity.

Alternatively, a defect in glycosylation might interfere with the proper processing of the interferon molecule by the producing cell in a way similar to that observed in cells infected with fowl plague virus. In the latter system, D-glucosamine and 2-deoxy-D-glucose were shown to inhibit viral hemagglutinin and neuraminidase synthesis

while
carb
pear
viral

W
M. L
This
Heal
NO1
and
cal
GM-

BEKI

D-
tic
C/

Cou
M

he
44

DOR
fe

A
EYL

P
GAN

D
ta

HAV
h

c
HUI

I
g

KAI
(

F
t

KA
I

v
(

KE
s

i
i

KI
:

KI

INTERFERON AND INHIBITORS OF GLYCOSYLATION

483

t the sugar determining the specificity of the action of target cells, that the two are exclusive.

2-deoxy-D-glucose interferon production process. Levels of interference, however, the rule out the per glycosylation. The release of ls.

Why proper for the synthesis of interferon. Although to speculate is important, interferon with a target cells, this is only in view of that rabbit fibroblastic activity (Schonne *et al.*, 1972) and enzymatic cleavage of carbohydrate chains in an interferon active (C. Schonne *et al.*, 1972). Although only in rate, it seems that the carbohydrate portion with a specificity of this the exposed cells. It is possible that the core chains may be a stability of which, in turn, biological activity.

glycosylation per processing the producing at observed in the virus. In the 2-deoxy-D-glucose inhibit viral hemagglutination synthesis

while leading to the accumulation of a carbohydrate-free polypeptide that appeared to be an inactive precursor of the viral glycoproteins (Klenk *et al.*, 1972).

ACKNOWLEDGMENTS

We thank Fermina Varacalli, Angel Feliciano, and M. Lyndle Gradoville for skilled technical assistance. This investigation was supported by U.S. Public Health Service Grant No. AI-07057 and Contract No. NO1-AI-02169 from the National Institute of Allergy and Infectious Diseases. B. B. is supported by Medical Scientist Training Program Grant No. 2-TO5-GM-01668.

REFERENCES

- BEKESI, J. G., and WINZLER, R. J. (1969). The effect of D-glucosamine on the adenine and uridine nucleotides of sarcoma 180 ascites tumor cells. *J. Biol. Chem.* **244**, 5663-5668.
- COURTNEY, R. J., STEINER, S. M., and BENYSH-MELNICK, M. (1973). Effects of 2-deoxy-D-glucose on herpes simplex virus replication. *Virology* **52**, 447-455.
- DORNER, F., SCRIBA, M., and WEIL, R. (1973). Interferon: Evidence for glycoprotein nature. *Proc. Nat. Acad. Sci. USA* **70**, 1981-1985.
- ENLAR, E. H. (1966). On the biological role of glycoproteins. *J. Theoret. Biol.* **10**, 89-113.
- GANDHI, S. S., STANLEY, P., TAYLOR, J. M., and WHITE, D. O. (1972). Inhibition of influenza viral glycoprotein synthesis by sugars. *Microbios* **5**, 41-50.
- HAVELL, E. A., and VILČEK, J. (1972). Production of high-titered interferon in cultures of human diploid cells. *Antimicrob. Ag. Chemother.* **2**, 476-484.
- HUNTER, E., FRIIS, R. R., and VOGT, P. K. (1974). Inhibition of avian sarcoma virus replication by glucosamine. *Virology* **58**, 449-456.
- KALUZA, G., SCHMIDT, M. F. G., and SCHOLTISSEK, C. (1973). Effect of 2-deoxy-D-glucose on the multiplication of Semliki Forest virus and the reversal of the block by mannose. *Virology* **54**, 179-189.
- KALUZA, G., SCHOLTISSEK, C., and ROTT, R. (1972). Inhibition of the multiplication of enveloped RNA-viruses by glucosamine and 2-deoxy-D-glucose. *J. Gen. Virol.* **14**, 251-259.
- KEPPLER, D. O. R., RUDIGER, J. F. M., BISCHOFF, E., and DECKER, K. F. A. (1970). The trapping of uridine phosphates by D-galactosamine, D-glucosamine, and 2-deoxy-D-galactose. *Eur. J. Biochem.* **17**, 246-253.
- KILBOURNE, E. D. (1959). Inhibition of influenza virus multiplication with a glucose antimetabolite (2-deoxy-D-glucose). *Nature (London)* **183**, 271-272.
- KLENK, H. D., SCHOLTISSEK, C., and ROTT, R. (1972). Inhibition of glycoprotein biosynthesis of influenza virus by D-glucosamine and 2-deoxy-D-glucose. *Virology* **49**, 723-734.
- LETNANSKY, K. (1964). The influence of 2-deoxy-D-glucose on the nucleotide content of Ehrlich ascites carcinoma cells. *Biochim. Biophys. Acta* **87**, 1-8.
- MARSHALL, L. W., PITHA, P. M., and CARTER, W. A. (1972). Inactivation of interferon: The effect of protonation. *Virology* **48**, 607-611.
- NG, M. H., BERMAN, B., and VILČEK, J. (1972). Membrane-bound intracellular interferon in rabbit kidney cells. *Virology* **49**, 322-325.
- SAMSON, A. C. R., and FOX, F. (1974). Selective inhibition of Newcastle disease virus-induced glycoprotein synthesis by D-glucosamine hydrochloride. *J. Virol.* **13**, 775-779.
- SCHOLTISSEK, C. (1971). Detection of an unstable RNA in chick fibroblasts after reduction of the UTP-pool by glucosamine. *Eur. J. Biochem.* **24**, 358-365.
- SCHOLTISSEK, C., ROTT, R., HAU, G., and KALUZA, G. (1974). Inhibition of the multiplication of vesicular stomatitis and Newcastle disease virus by 2-deoxy-D-glucose. *J. Virol.* **13**, 1186-1193.
- SCHONNE, E., BILLIAU, A., and DESOMER, P. (1970). The properties of interferon. IV. Isoelectric focusing of rabbit interferon (NDV-RK 13). In "International Symposium on Standardization of Interferon and Interferon Inducers" (F. T. Perkins and R. H. Regamey, eds.), pp. 61-68. Karger, Basel, New York.
- SPIRO, R. G. (1970). Glycoproteins. *Annu. Rev. Biochem.* **39**, 599-638.
- SPIRO, R. G. (1973). Glycoproteins. *Advan. Protein Chem.* **27**, 349-467.
- TAN, Y. H., ARMSTRONG, J. A., and HO, M. (1971). Intracellular interferon: Kinetics of formation and release. *Virology* **45**, 837-840.
- TAN, Y. H., JENG, D. K., and HO, M. (1972). The release of interferon: An active process inhibited by p-hydroxymercuribenzoate. *Virology* **48**, 41-48.
- VILČEK, J. (1970). Metabolic determinants of the induction of interferon by a synthetic double-stranded polynucleotide in rabbit kidney cells. *Ann. N.Y. Acad. Sci.* **173**, 390-403.
- VILČEK, J., and HAVELL, E. A. (1973). Stabilization of interferon messenger RNA activity by treatment of cells with metabolic inhibitors and lowering of the incubation temperature. *Proc. Nat. Acad. Sci. USA* **70**, 3909-3913.
- VILČEK, J., and HAVELL, E. A. (1974). Interferon production as a model of induced protein synthesis in eukaryotic cells. In "Proceedings of the Gulbenkian Meeting on Interferon, Held in Oeiras, Sept. 19-21, 1973" (in press).
- WEIL, R., and DORNER, F. (1973). Interferon structure: facts and speculation. In "Selective Inhibitors of Viral Functions" (W. A. Carter, ed.), pp. 107-121. CRC Press, Cleveland, OH.
- WINTERBURN, P. J., and PHELPS, C. F. (1972). The significance of glycosylated proteins. *Nature (London)* **236**, 147-151.

TAB XX

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 251, No. 6, Issue of March 25, pp. 1659-1662, 1976
Printed in U.S.A.

Apparent Dispensability of the Carbohydrate Moiety of Human Interferon for Antiviral Activity

(Received for publication, August 18, 1975)

SIKTA BOSE, DALIA GURARI-ROTMAN, URS TH. RUEGG,* LILA CORLEY, AND CHRISTIAN B. ANFINSEN

From the Laboratory of Chemical Biology, National Institute of Arthritis, Metabolism, and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20014

Human leukocyte and tritium-labeled fibroblast interferons, prepared by induction with Sendai virus and with double-stranded polyinosinic acid-polycytidylic acid respectively, have been studied in relation to the carbohydrate moieties attached to them. These interferons were partially purified by immunoabsorbance and by gel filtration. On treatment with glycosidases, about 80% of the ^3H -labeled sugar moieties in this glycoprotein-containing fraction was removed without detectable alteration of the antiviral activity or antibody-binding properties characteristic of interferon. The molecular weight of leukocyte interferon was reduced by about 4000. As others have reported, the heterogeneous character of interferon revealed by isoelectric focusing was greatly reduced by the enzyme treatment.

The study of the purification and properties of interferon has attracted much attention because of the potentially valuable antiviral activity of this substance. It has not yet been possible to produce interferon in large quantities by tissue culture. Consequently, many investigators have focused either on its isolation in pure form, for the purpose of determination of structure and, possibly, for future chemical synthesis, or on the development of potent, nontoxic inducers to stimulate endogenous interferon production.

Schonne *et al.* (1) observed a decreased charge heterogeneity of rabbit interferon upon isoelectric focusing following treatment with neuraminidase. They proposed that interferon is a glycoprotein, possessing variable amounts of sialic acid. Later, this hypothesis was confirmed by Dorner *et al.* (2). Rabbit interferon, upon treatment with neuraminidase, exhibited a single peak focusing at pH 6.3. This asialo-interferon regained its original heterogeneous character after reincorporation of sialic acid residues by the use of sialyltransferase. Also, the terminal sequence of the carbohydrate moiety was identified as sialic acid — galactose.

The studies of Schonne *et al.* (1) and of Dorner *et al.* (2), referred to above, have shown that terminal sialic residues may be removed without loss of antiviral activity. In the case of the latter studies (2), the next carbohydrate residue in the polysaccharide chain, galactose, could be successively oxidized with galactose oxidase and reduced with sodium borohydride without loss of activity. Since most glycoproteins studied to date contain one or more polysaccharide chains with molecular weights of 1000 to 3000, the present studies were carried out to examine the effects of more extensive carbohydrate removal.

Such studies are of importance in connection with the feasibility of eventual synthesis. This paper summarizes experiments on human leukocyte and fibroblast interferon that

indicate the dispensability of the bulk of the carbohydrate component.

MATERIALS AND METHODS

Human leukocyte interferon prepared from human leukocytes induced with Sendai virus by H. Strander and K. Cantell (3) and Smith, Kline and French Laboratories was supplied to us by the Antiviral Substances Program of the National Institute for Allergy and Infectious Diseases.

Fibroblast interferon was prepared with the kind cooperation of J. Vilček and E. Havell at New York University, using the FS-4 strain of human foreskin diploid fibroblasts induced with a double-stranded complex of polyinosinic acid-polycytidylic acid (4). Tritiated glycosides were incorporated into interferon preparations by adding D- ^3H glucosamine-hydrochloride with a specific activity of 10.13 Ci/mM (New England Nuclear, 5 mCi/100 ml of culture fluid) to Eagle's minimal essential medium with a glucose concentration lowered to 20 mg/100 ml. The tissue culture fluids were first desalted by passing them through a column of Sephadex G-25 medium (5.2 × 75 cm) in 0.1 M acetic acid to remove the bulk of the radioactivity.

The ^3H -labeled protein fraction from fibroblast cultures was purified by immunoabsorbance chromatography (5, 6) and further by gel filtration on Sephadex G-100 (fine 2.6 × 90 cm). Human leukocyte interferon was subjected to the same purification procedures. Interferons were eluted from the affinity column by a pH-gradient (6) containing cytochrome c (0.5 mg/ml) and from G-100 columns by 0.1 M acetic acid containing NaCl (0.15 M) and cytochrome c (0.5 mg/ml). The partially purified interferons were dialyzed against 0.1 M acetic acid and lyophilized.

Interferon assays were carried out by a modification of the micro method originally described for rabbit kidney cells (7). The reference human interferon preparation, 69/19, titers 5×10^3 units/ml in this assay. The specific activities of partially purified interferons have not been estimated in the current studies. As discussed earlier (6), a large purification (200 to 300-fold) is obtained in the first step (immunoabsorbance on antibody-Sepharose columns). Indeed, the amounts of protein present in the interferon fraction is sufficiently low that it is necessary to add a protective protein (in our experiments, cytochrome c) to prevent losses due to adsorption in subsequent steps. The current studies are concerned with the physical behavior before and after removal of carbohydrates. The changes in charge and mass have been

related to the migration of material possessing antiviral activity during isoelectric focusing and gel filtration.

The mixture of glycosidases used to partially remove the carbohydrate moiety from interferon was prepared (8) from a strain of *Diplococcus pneumoniae*, Type 1, with the kind cooperation of G. Ashwell, National Institute of Arthritis, Metabolism, and Digestive Diseases. α -Mannosidase (49 units/ml) was a gift from Dr. Ashwell. The glycosidase preparation (8, 9) contained β -galactosidase (4.6 units/ml), *N*-acetylglucosaminidase (10.4 units/ml), neuraminidase (2.1 units/ml), and an undetermined amount of endoglycosidase (10). The protein concentration of this preparation was 3.9 mg/ml.

Treatment of ^3H -labeled Interferon with Glycosidases—The partially purified and lyophilized interferon was redissolved in McIlvaine phosphate/citrate buffer, pH 6.0, and the initial radioactivity was determined. Solutions of glycosidases (2 μl /1000 units of interferon) and α -mannosidase (0.25 μl /1000 units of interferon) were added and the mixture was incubated at 37°. Aliquots (0.2 ml) were taken during the course of the enzymic reaction for interferon assay. For the determination of sugar released, aliquots (0.1 ml) were removed and precipitated by addition of 0.2 ml of 2% phosphotungstic acid in 0.5 *N* HCl. After centrifugation, a measured amount of the supernatant was counted in a liquid scintillation counter (Table I). (Leukocyte interferon was also treated with glycosidases and α -mannosidase under the same conditions as described above; see Figs. 1 and 2.)

For experiments designed to identify the ^3H -labeled sugars removed by enzyme treatment, the interferon was dissolved in deionized water rather than buffer. After 3 hours incubation with glycosidases alone (2 μl /1000 units of interferon), the reaction mixture was passed through a column (1.6 \times 28 cm) of Sephadex G-25 medium equilibrated with 0.1 *M* acetic acid. The fractions containing the low molecular weight material were lyophilized and redissolved in 0.5 ml of water. A 0.1-ml aliquot was applied together with a carrier mixture of mannose, galactose, sialic acid, and *N*-acetylglucosamine (1 μg each) on a Whatman No. 1MM paper. The same mixture was also applied for reference and the paper was subjected to descending chromatography using as solvent system, pyridine/acetic acid/ethyl acetate/water (5/1/5/3, v/v). The reference strip was developed with silver nitrate-solution (11) and the corresponding areas of the strip with the radioactive sample were cut out and counted in a liquid scintillation counter.

Antibody-binding Experiments—Antibodies to human leukocyte interferon, prepared in sheep, were purified and bound to Sepharose as described previously (6). Leukocyte and fibroblast interferons, after treatment with enzymes, were passed through a column (1 \times 5 cm) prepared from the Sepharose-bound antibodies. The column was washed with phosphate-buffered saline and eluted at pH 2.2 with Sorensen's citrate buffer containing cytochrome c (0.5 mg/ml).

Isoelectric Focusing—Leukocyte interferon (6.7×10^4 units), before and after treatment with glycosidases and α -mannosidase for 3 hours as mentioned above for fibroblast material, was subjected to isoelectric focusing at 4° using pH 3.5 to 8.0 Ampholines (LKB, Bromma, Sweden) and a 25-ml sucrose gradient of 0 to 50% (Fig. 1). After 5 hours at 310 volts, 0.5-ml fractions were collected and their pH was measured. Fractions were assayed for interferon activity.

Molecular Weight Determination—The apparent molecular weight of human leukocyte interferon was estimated by gel filtration using a Sephadex G-100 (fine) column (1.5 \times 100 cm) equilibrated with 0.1 *M* acetic acid containing 0.15 *M* NaCl. Ovalbumin, chymotrypsinogen, and bovine pancreatic ribonuclease were added as markers and dextran blue was used to determine the void volume (elution followed at 280 nm on LKB Uvicord). The flow rate was maintained at 10 ml/hour and 1-ml fractions were collected. The eluates were examined at 280 nm for protein concentration, and assayed for interferon and ribonuclease (12) activities. The determination was carried out on leukocyte interferon (10^4 units) both before and after treatment with glycosidases and α -mannosidase.

RESULTS AND DISCUSSION

As can be seen in Table I, incubation of the protein fraction containing fibroblast interferon labeled with tritiated sugar moieties with a mixture of glycosidases and α -mannosidase results in the progressive removal of tritiated sugar residues. Between 74 to 85% of the tritium label was removed without loss of interferon activity in three experiments carried out on different preparations of tritiated interferon. Complete re-

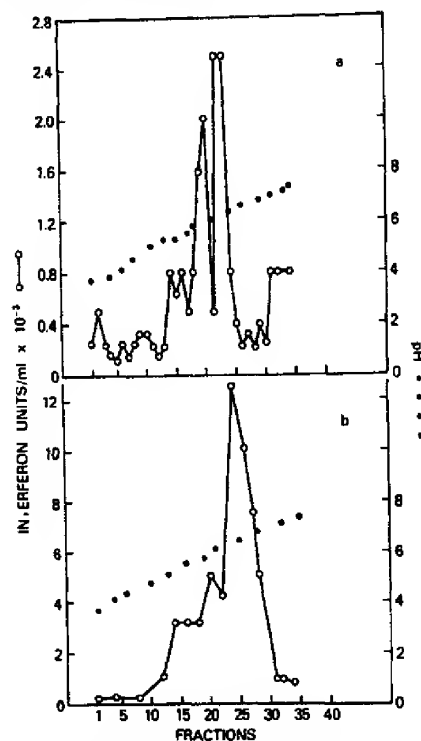


FIG. 1. Isoelectric focusing of human interferon before (a) and after (b) removal of the carbohydrate moiety by treatment with glycosidases for 3 hours as described under "Materials and Methods." Human interferon (leukocyte, 6.7×10^4 units in each case) was focused at pH 3.5 to 8 in a 0 to 50% sucrose gradient. ●, pH profile; ○—○, interferon activity.

moval of the tritiated sugars was not achieved, even with longer enzyme incubation times. This could be due to the fact that part of the *N*-acetylglucosamine residues are directly connected to the polypeptide chain of the glycoprotein and that the sugar-amino acid bond cannot be enzymatically cleaved. However, the fact that the molecular weight (Fig. 2) of leukocyte interferon is considerably decreased by glycosidase treatment indicates that a large part of the sugar attached to interferon has been removed.

The tritiated sugars cleaved from the interferon-containing glycoprotein fraction by treatment with glycosidases alone were separated by paper chromatography. Of the total radioactivity applied to the paper (572 dpm), 124 dpm were recovered in the spot corresponding to sialic acid ($R_f = 0.50$) and 348 dpm as *N*-acetylglucosamine ($R_f = 0.69$). No significant radioactivity was detected in any other part of the paper strip. Thus, some conversion of the glucosamine, added to the culture medium during interferon production, to sialic acid by condensation with phosphoenolpyruvate and to *N*-acetylglucosamine by reaction with acetyl-CoA had taken place. Conversion of glucosamine to other sugars commonly occurring in glycoproteins would not be expected under our experimental conditions.

Treatment of fibroblast interferon, prepared in the presence of tritiated glucosamine, with neuraminidase alone resulted in the removal of 25 to 30% of the tritium label and in this case the only product detected was [^3H]sialic acid. This agrees with the findings of Dorner *et al.* (2) for rabbit interferon, namely

TABLE I
Enzymic cleavage of tritiated sugar moieties from human fibroblast interferon

| Interferon preparation ^a | Incubation time min | Amount of ³ H label in supernatant | ³ H label released | Interferon activity units/ml × 10 ⁻³ |
|-------------------------------------|------------------------|---|-------------------------------|--|
| | | dpm/ml × 10 ⁻³ | % | |
| I ^b | 0 | 0.44 | 1.2 | 25 |
| | 10 | 5.68 | 16.3 | |
| | 20 | 10.96 | 31.5 | 16 |
| | 40 | 17.98 | 51.7 | |
| | 60 | 20.26 | 58.3 | 25 |
| | 80 | 23.42 | 67.4 | |
| | 120 | 25.88 | 74.5 | 25 |
| | 160 | 27.38 | 78.8 | |
| | 240 | 28.07 | 80.7 | 25 |
| II | 0 | 0.3 | 0.1 | 3.2 |
| | 180 | 392 | 73.4 | 3.2 |
| III | 0 | 0.2 | 0.2 | 5.0 |
| | 180 | 94.6 | 84.7 | 5.0 |

^a Interferon Preparations I, II, and III contained 3.48×10^5 , 5.34×10^5 , and 1.12×10^5 dpm/ml, respectively.

^b In this experiment, a 3-fold excess of unlabeled fibroblast interferon was added to the labeled material.

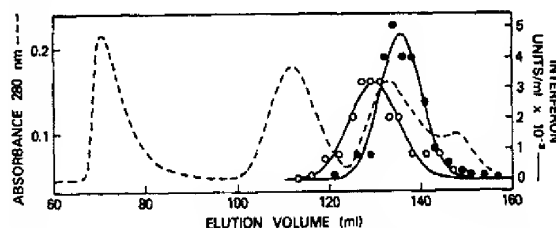


FIG. 2. Apparent molecular weight determination by Sephadex (G-100, fine) gel filtration of human interferon before and after cleavage of the sugar moiety. The column (1.5 × 100 cm), equilibrated and eluted with 0.1 M acetic acid/0.15 M NaCl, was used for the separation of: first, dextran blue, ovalbumin, chymotrypsinogen, and ribonuclease; second, ovalbumin, ribonuclease, and interferon (leukocyte; 10^4 units); third, ovalbumin, ribonuclease, and interferon treated with glycosidases and α -mannosidase as described under "Materials and Methods." The A_{280} elution profiles of the marker proteins of the three column runs were completely indistinguishable when they were superimposed; 1-ml fractions were collected. --- ultraviolet profile at 280 nm of markers; —, interferon activity of (O) untreated and (●) enzyme-treated interferon. The experimentally determined values for interferon activity were fitted by Gaussian curves, assuming each point to have the same statistical value.

that sialic acid is present as the terminal residue in the carbohydrate moiety of this material.

Neither leukocyte nor fibroblast interferons showed any change in binding to the anti-interferon antibody column as a result of treatment with glycosidases and α -mannosidase (Table II) under the conditions described under "Materials and Methods." In one experiment (No. 1B, Table II), the enzyme-treated interferon solution (4.4×10^4 dpm) was applied to this affinity column. Most of the radioactivity (3.6×10^4 dpm; 81%) passed through the column unretarded, while 0.84×10^4 dpm (19%) were eluted with citrate buffer (pH 2.2) along with the interferon activity. These figures are in agreement with those from the experiments described in Table I.

TABLE II
Immunoabsorbance chromatography of human leukocyte and fibroblast interferons before and after treatment with glycosidases and α -mannosidase

| Experiment No. | Type of interferon | Enzyme treatment | Total units applied, $\times 10^{-3}$ | Units bound to anti-interferon column, $\times 10^{-3}$ | Recovery ^a % |
|----------------|--|------------------|---------------------------------------|---|-------------------------|
| 1A | Fibroblast | — | 19 | 8.5 | 45 |
| 1B | Fibroblast, ³ H-labeled sugar | + | 9.9 | 4.3 | 43 |
| 2A | Leukocyte | — | 90 | 53 | 59 |
| 2B | Leukocyte | + | 90 | 64 | 71 |

^a Recovery refers to the antibody-bound interferon, eluted with citrate buffer, pH 2.2.

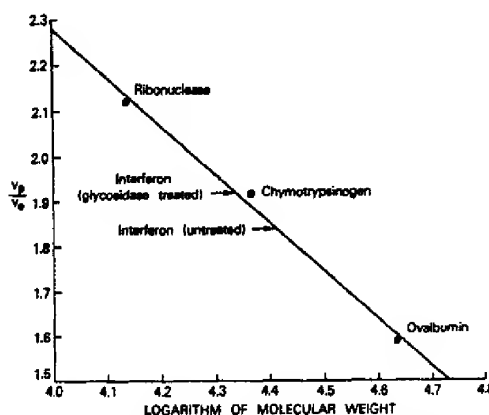


FIG. 3. Graphic determination of the molecular weight of interferons before and after enzymic removal of the carbohydrate portions (Fig. 2). The ratios of the elution volume of each protein (V_p) divided by the elution volume of blue dextran (V_o) were plotted against the logarithm of molecular weights of the protein markers: ovalbumin (4.3×10^4), chymotrypsinogen (2.32×10^4), and ribonuclease (1.37×10^4).

This indicates that the removal of much of the carbohydrate from interferon does not affect its ability to bind to its specific antibodies.

Isoelectric focusing experiments on both types of interferons indicate that after partial removal of the carbohydrate moiety from interferon, its heterogeneous character is reduced (Fig. 1). The recovery of leukocyte interferon (6.7×10^4 units) on isoelectric focusing was 36%, whereas the recovery after enzymic removal of the carbohydrate moiety was 71%. This type of heterogeneous behavior on isoelectric focusing was previously observed by Schonke *et al.* (1) using rabbit interferon and by Stanček *et al.* (13) using rabbit, mouse, and human interferon. The heterogeneity was reduced after treatment with neuraminidase (1, 2). Dorner *et al.* (2) showed that the terminal sugar residue of the carbohydrate moiety was sialic acid.

The results of molecular weight determinations on leukocyte interferon before and after treatment with glycosidases are shown in Figs. 2 and 3. After enzyme treatment (glycosidases and α -mannosidase), the apparent molecular weight of the interferon is decreased by about 4,200. This observation was verified in three separate experiments. The molecular weight

determined for untreated interferon is $26,000 \pm 1,200$. It should be noted, however, that glycoproteins are not believed to behave as globular proteins on gel filtration (14). In addition, since losses of interferon activity occurred during gel filtration in neutral buffers, our columns were eluted under acid conditions which may cause some protein unfolding. Therefore, we stress that the change in molecular weight of interferon is the important observation rather than the absolute molecular weight values.

Acknowledgments—We are grateful to Doctors E. Havell and J. Vilček for their help in the preparation of fibroblast interferon, to Dr. G. Ashwell for the gift of α -mannosidase as well as his help in the preparation of glycosidases and to Dr. A. Minton, National Institute of Arthritis, Metabolism, and Digestive Diseases, for computer processing of the molecular weight data.

REFERENCES

1. Schonne, E., Billiau, A., and DeSomer, P. (1970) in *International Symposium on Standardization of Interferon and Interferon Inducers* (Perkins, F. T., and Regamey, R. H., eds) Vol. 14, pp. 61-68, Karger, Basel
2. Dorner, F., Scriba, M., and Weil, R. (1973) *Proc. Natl. Acad. Sci. U. S. A.* **70**, 1981-1985
3. Strander, H., and Cantell, K. (1966) *Ann. Med. Exp. Biol. Fenn.* **44**, 265-273
4. Vilček, J., and Havell, E. A. (1973) *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3909-3913
5. Campbell, D. H., Leuscher, E., and Lermann, L. S. (1951) *Proc. Natl. Acad. Sci. U. S. A.* **37**, 575-578
6. Anfinsen, C. B., Bose, S., Corley, L., and Gurari-Rotman, D. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71**, 3139-3142
7. Armstrong, J. A. (1971) *Appl. Microbiol.* **21**, 723-725
8. Hughes, R. C., and Jeanloz, R. W. (1964) *Biochemistry* **3**, 1535-1543
9. Warren, L. (1959) *J. Biol. Chem.* **234**, 1971-1973
10. Koide, N., and Muramatsu, T. (1974) *J. Biol. Chem.* **249**, 4897-4904
11. Trevelyan, W. E., Procter, D. P., and Harrison, J. S. (1950) *Nature* **166**, 444-445
12. Crook, E. M., Mathias, A. P., and Rabin, B. R. (1960) *Biochem. J.* **74**, 234-238
13. Stanček, D., Gressnerova, M., and Paucker, K. (1970) *Virology* **41**, 740-750
14. Andrews, P. (1965) *Biochem. J.* **96**, 595-606

TAB YY

Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 252, No. 12, Issue of June 26, pp. 4425-4427, 1977
Printed in U.S.A.

Altered Molecular Species of Human Interferon Produced in the Presence of Inhibitors of Glycosylation*

(Received for publication, March 30, 1977)

EDWARD A. HAVELL, SHUDO YAMAZAKI,† AND JAN VILČEK
From New York University School of Medicine, Department
of Microbiology, New York, New York 10016

SUMMARY

The inhibitors of glycosylation, 2-deoxy-D-glucose or D-glucosamine, inhibit the synthesis of biologically active interferon in human FS-4 fibroblast cultures stimulated with polyinosinate·polycytidylylate. Interferon synthesized in the presence of partially inhibitory concentrations of 2-deoxy-D-glucose or D-glucosamine were found to differ from interferons made in control cultures in some physical properties. Interferons synthesized in the presence of either inhibitor had a diminished charge heterogeneity demonstrable by isoelectric focusing. On sodium dodecyl sulfate-polyacrylamide gel electrophoresis, control interferon activity formed a single peak with the apparent molecular weight of 20,000, whereas interferons from cultures treated with either inhibitor could be resolved into two distinct molecular weight components, one of which was smaller than the interferon synthesized in control cultures.

Most interferons are probably glycoproteins. Two groups of investigators (1, 2) observed that neuraminidase treatment of rabbit interferon resulted in a considerable decrease in the charge heterogeneity observed on isoelectric focusing. Dorner *et al.* (2) were able to restore the charge heterogeneity of desialylated rabbit interferon by its incubation with *N*-acetylneuraminic acid and sialyltransferase.

Several lines of evidence suggest that human interferon derived from cultures of skin fibroblast cells is also glycosylated. Jankowski *et al.* (3) have shown that human fibroblast interferon was capable of binding to several immobilized lectins specific for *N*-acetylneuraminic acid, L-fucose, and D-mannose. Studies on the intracellular site of human fibroblast interferon synthesis have shown that it is probably synthesized in the membranes of the rough endoplasmic reticulum (4) and that it may follow the secretory pathway of other known glycoproteins (5, 6). The synthesis of interferon is inhibited in a dose-dependent manner by the inhibitors of glycosylation, 2-deoxy-D-glucose, and D-glucosamine (7). The 2-deoxy-D-glucose is known to act as an anti-metabolite for mannose (8), whereas the precise mechanism of action of D-glucosamine remains unknown (9).

We have extended our studies on the physical nature of the interferons synthesized in the presence of partially inhibitory concentrations of either 2-deoxy-D-glucose or D-glucosamine.

* This work was supported by United States Public Health Service Grants AI-07057, AI-12948, and Contract AI-02169 from the National Institute of Allergy and Infectious Diseases.

† Permanent address, National Institute of Health of Japan, Murayama Annex, Nakato, Musashimurayama, Tokyo, 190-12, Japan.

In this communication, we report on the effects of these inhibitors of glycosylation on the molecular weight and isoelectric point of the interferons synthesized in cultured human fibroblast cells.

MATERIALS AND METHODS

Cultures of the human FS-4 fibroblast strain were grown in 1/2-gallon roller bottles and induced with the synthetic polyribonucleotide poly(I)·poly(C),¹ according to the method described earlier (10). Cultures were incubated for 6 h with poly(I)·poly(C) (5 µg/ml, P-L Biochemicals, Milwaukee, Wis.) together with cycloheximide (50 µg/ml, Upjohn Co., Kalamazoo, Mich.) in 50 ml of Eagle's Minimal Essential Medium (MEM, GIBCO). Actinomycin D (1 µg/ml, Calbiochem, San Diego, Calif.) was added to the induction medium at 5 h. At the end of the 6-h induction period the medium was removed, and the cultures extensively washed and then reincubated at 34° for an additional 24 h in the interferon production medium which consisted of 50 ml of MEM buffered with 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Hepes, 13 mM), *N*-[Tris(hydroxymethyl)methyl]glycine (Tricine, 6 mM), and sodium bicarbonate (1.1 g/liter), supplemented with 0.5% heated fetal calf serum (56° for 30 min). Glucosamine or 2-deoxy-D-glucose was added to the production media at a final concentration of 10 mM and 5 mM, respectively. These concentrations of glucosamine or 2-deoxy-D-glucose inhibited the production of biologically active interferon by about 90%, as compared to inhibitor-free control cultures.

The interferon activities of the preparations were determined by the microtiter assay method as previously reported (8). The interferon preparations were concentrated approximately 10-fold by dialysis against the hygroscopic agent, Ficoll 400, prior to use in these experiments. Following concentration, the interferon titers for the control, 2-deoxy-D-glucose, and glucosamine interferons were 200,000, 25,000, and 6,000 units/ml, respectively. The lower titers of the preparations from inhibitor-treated cultures are due to the inhibitory action of 2-deoxy-D-glucose and glucosamine on interferon synthesis (7).

RESULTS AND DISCUSSION

The concentrated interferon preparations obtained from induced cultures of FS-4 fibroblast cells treated with or without 5 mM 2-deoxy-D-glucose were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (11, 12) to determine the molecular weights of each interferon species present in either preparation (Fig. 1). The interferon activity found in the interferon preparations obtained from the untreated FS-4 cultures (Control A) migrated as a homogeneous species with a molecular weight of 20,000. This molecular weight agrees with the value of Knight (13) for purified FS-4 fibroblast interferon and with our own earlier reported results (14). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of the interferon preparations from induced FS-4 cultures treated with 2-deoxy-D-glucose or glucosamine revealed the presence of two distinct interferon peaks, with about equal distribution of interferon activity in the two peaks. The slower migrating interferon species possessed a molecular weight similar to that found for the control interferon, whereas the faster migrating interferon activity was determined to have a molecular weight of 16,000. It should be noted that the lower activities for the interferons produced in the presence of either inhibitor of glycosylation are due to smaller amounts of interferon placed on the sodium dodecyl sulfate gels and not to the less efficient

¹ The abbreviation used is: poly(I)·poly(C), polyinosinate·polycytidylylate.

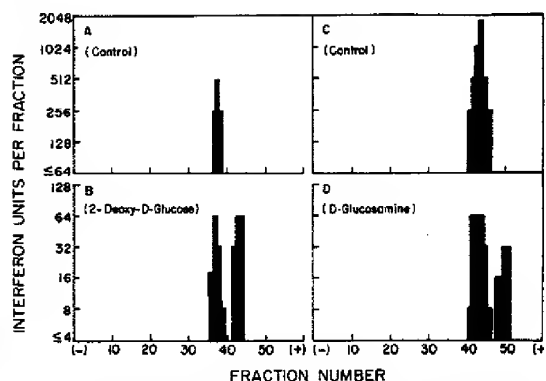


FIG. 1. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of interferons made in the presence of either 2-deoxy-D-glucose or D-glucosamine. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was done according to the method of Stewart and Desmyter (12). This electrophoresis procedure uses 10% polyacrylamide and 0.1 M sodium phosphate buffer, pH 7.4, with 0.1% sodium dodecyl sulfate. Prior to electrophoresis, the samples were incubated at 37° for 1 h in 1% sodium dodecyl sulfate, then dialyzed in 0.01 M sodium phosphate buffer, pH 7.4, with 0.1% sodium dodecyl sulfate. The dialyzed samples were then made to 10% in sucrose and electrophoresed for 18 h. The current applied was 5 mA/disc gel. Lysozyme, chymotrypsin, ovalbumin, and bovine serum albumin were used as markers for molecular weight determinations. Following electrophoresis, the gels were frozen and sliced into 1-mm slices, 2 adjacent slices (referred to as one fraction) were pooled, and interferon activity was eluted in 1 ml of MEM containing 5% fetal calf serum at room temperature overnight. Approximately 50% of the interferon activity applied to the sodium dodecyl sulfate gels was recovered in all four groups.

recovery of these interferon species from the sodium dodecyl sulfate gels.

The resulting isoelectric focusing distribution of interferon activities for the interferons from untreated and inhibitor-treated cultures is shown in Fig. 2. The interferon activity present in the control preparation was recovered from two regions of the pH gradient. Considerable amounts of interferon activity were recovered from the top of the gels (pH 4.0). This portion of the interferon activity may either not have entered the gel, or it may represent interferon noncovalently bound to albumin (16) or other proteins with an isoelectric point near pH 4.0. The rest of interferon activity was recovered from a very broad region centering just past neutrality, with the bulk of interferon activity focusing in the pH region 7.2 to 7.6. This broad heterodisperse activity focusing in the range of neutrality was not apparent in either of the two interferon preparations prepared in the presence of either 2-deoxy-D-glucose or D-glucosamine. Instead, both of these interferon preparations focused rather sharply, with most of the interferon activity found in a single fraction corresponding to a pH of 7.4. Considerable differences were noticed in the isoelectric focusing patterns of interferon preparations from the two inhibitor-treated FS-4 cultures in the acidic regions of the pH gradient. It is of interest that, compared to control interferon, less of the 2-deoxy-D-glucose interferon, and even less of the interferon made in the presence of D-glucosamine, was recovered from the top of the gradients. Although this result is not completely understood, it may indicate that aggregation of interferon with albumin or other material is the function of the carbohydrate moiety.

These results supplement our previous observations on the inhibitory action of 2-deoxy-D-glucose and D-glucosamine on

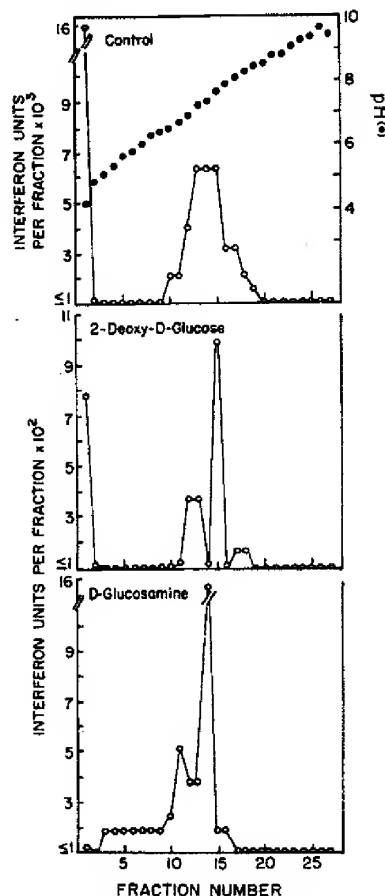


FIG. 2. Isoelectric focusing of interferons synthesized in the presence of either 2-deoxy-D-glucose or D-glucosamine. A modification of the isoelectric focusing procedure of Gainer (15) was used to determine the isoelectric points (pIs) of the interferons. Cylindrical disc gels (0.6 x 8 cm) consisted of 1.8 ml of 7.5% polyacrylamide containing 2% ampholytes (LKB, Hicksville, New York, pH range 3.5 to 10). Samples to be isoelectrically focused were dialyzed against a 1% glycine-distilled H₂O solution, made to 10% in sucrose, and 0.2 ml of this material was layered on the anode end of the gel. The solutions in the electrode vessels were 1% H₂SO₄ (anode) and 2% ethanolamine (cathode). Isoelectric focusing was done at 4° over approximately a 5-h period, with an initial potential of 150 V. Fractionation and elution of interferon was essentially as described for sodium dodecyl sulfate-polyacrylamide gels, except elution was done at 4°. Approximately 25% of the interferon activity applied to the gels containing ampholytes was recovered in all groups. Determination of the pH gradient was done by eluting fractions of a blank gel overnight at 4° in distilled H₂O and measuring the pH at 4°.

the synthesis of human fibroblast interferon (7). They support more directly the conclusion that the interferon molecules synthesized in the presence of inhibitors are altered. Our sodium dodecyl sulfate-polyacrylamide gel electrophoresis results suggest that interferons made in the presence of partially inhibitory concentrations of the two drugs may be a mixture of altered and unaltered (or minimally altered) interferon molecules. If the differences in the molecular weights of the two interferon peaks are determined by the presence or absence of carbohydrates, then the carbohydrate would be likely to account for at least 20% of the total molecular weight of human fibroblast interferon (16,000 versus 20,000 molecular weight).

Glycosylation of Human Interferon

4427

This calculation is in good agreement with an independent determination obtained by incubating human interferon with various purified glycosidases (17). This treatment resulted in a decrease in molecular weight and reduction of the charge heterogeneity of the interferon preparation.

Acknowledgments—We would like to acknowledge the excellent technical assistance of M. Lyndle Gradoville and Angel Feliciano.

REFERENCES

- Schonne, E., Billiau, A., and De Somer, P. (1970) in *International Symposium on Standardization of Interferon Inducers* (Perkins, F. T., and Regamey, R. H., eds) Vol. 14, pp. 61-68, Karger, Basel
- Dorner, F., Scriba, M., and Weil, R. (1973) *Proc. Natl. Acad. Sci. U. S. A.* 70, 1981-1985
- Jankowski, W. J., Davey, M. W., O'Malley, J. A., Sulkowski, E., and Carter, W. A. (1975) *J. Virol.* 16, 1124-1130
- Falcoff, E., Havell, E. A., Lewis, J. A., Lande, M. A., Falcoff, R., Sabatini, D. D., and Vilček, J. (1976) *Virology* 75, 384-393
- Caro, L. G., and Palade, G. E. (1964) *J. Cell Biol.* 20, 473-495
- Jamieson, J. D., and Palade, G. E. (1967) *J. Cell Biol.* 34, 577-596
- Havell, E. A., Vilček, J., Falcoff, E., and Berman, B. (1975) *Virology* 63, 475-483
- Kaluza, G., Schmidt, M. F. G., and Scholtissek, C. (1973) *Virology* 54, 179-189
- Scholtissek, C. (1975) *Curr. Top. Microbiol. Immunol.* 70, 101-119
- Havell, E. A., and Vilček, J. (1972) *Antimicrob. Agents Chemother.* 2, 476-484
- Sumers, D. F., Maizel, J. V., and Darnell, J. E. (1965) *Proc. Natl. Acad. Sci. U. S. A.* 54, 505-513
- Stewart, W. E., II, and Desmyter, J. (1975) *Virology* 67, 68-73
- Knight, E., Jr. (1976) *Proc. Natl. Acad. Sci. U. S. A.* 73, 520-523
- Vilček, J., Havell, E. A., and Yamazaki, S. (1977) *Ann. N. Y. Acad. Sci.* 284, 703-710
- Gainer, H. (1973) *Anal. Biochem.* 51, 646-650
- Huang, J. W., Davey, M. W., Hejna, C. J., von Muenchhausen, W., Sulkowski, E., and Carter, W. A. (1974) *J. Biol. Chem.* 249, 4665-4667
- Bose, S., Gurari-Rotman, D., Ruegg, U. Th., Corly, L., and Anfinsen, C. B. (1976) *J. Biol. Chem.* 251, 1659-1662

TAB ZZ

Expression of the human fibroblast interferon gene in *Escherichia coli**

(hybrid ribosome-binding site/antiviral activity/*lacZ* gene fusions/portable promoter)

TADATSUGU TANIGUCHI[†] AND
LEONARD GUARENTE[‡], THOMAS M. ROBERTS[‡], DAVID KIMELMAN[‡], JOHN DOUHAN III[‡], AND
MARK PTASHNE[‡]

[†]Cancer Institute, Japanese Foundation for Cancer Research, Toshima-Ku, Tokyo 170, Japan; and [‡]The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Contributed by Mark Ptashne, June 26, 1980

ABSTRACT We applied the method of Guarente *et al.* [Guarente, L., Lauer, C., Roberts, T. M. & Ptashne, M. (1980) *Cell* 20, 543-553] to construct plasmids that direct expression in *Escherichia coli* of the human fibroblast interferon (F-IF) gene. Two plasmids were recovered. One directs efficient synthesis of a protein whose primary sequence is that of pre-F-IF and the other, that of mature F-IF. Extracts of bacteria synthesizing mature F-IF display antiviral activity characteristic of human F-IF. This activity is lower than that expected from the differential rate of synthesis of the protein. We have detected no such activity in extracts of bacteria synthesizing pre-F-IF.

Human fibroblast interferon (F-IF) is a glycoprotein produced by human fibroblasts in response to virus and certain polynucleotides (1, 2). The secreted protein has potent antiviral activity that is readily assayed *in vitro* (2, 3). The sequence of the amino-terminal 13 amino acids of F-IF has been reported (4).

A cDNA molecule encoding human F-IF was cloned by Taniguchi *et al.* (5). The DNA sequence of this molecule predicts that the secreted F-IF contains 166 amino acids, the first 13 of which would be identical to the corresponding sequence of the protein as determined by Knight *et al.* (4). Moreover, the sequence is consistent with the idea that F-IF is synthesized as a precursor (pre-F-IF) with a 21-amino-acid hydrophobic leader at its amino terminus (6, 7).

A series of papers from this laboratory* has developed methods to express cloned prokaryotic and eukaryotic genes in *Escherichia coli* (8-11). The protein products of these plasmid-carried genes were produced in their native states—that is, unfused to other proteins (8-11). The essential feature of this method is to position a “portable promoter” in front of the cloned gene so that the gene is efficiently transcribed and the resultant mRNA is efficiently translated, beginning at an initiation codon—e.g., AUG. This AUG may, but need not be, that which directs initiation of synthesis of the native protein *in vitro*. The procedure of Roberts *et al.* (9, 10) enables us to position the portable promoter at various positions in front of the cloned gene by using recombination *in vitro*. The method of Guarente *et al.* (11) exploits *lac* genetics to identify those positionings that direct efficient transcription and translation of the cloned gene. This latter procedure eliminates the need for any assay for the gene product to identify those bacteria that express the desired proteins (see *Method of Gene Expression in Results*).

The protein and DNA sequence data referred to above indicate that both F-IF and pre-F-IF bear methionine residues

at their amino termini (4, 6). We describe in this paper the application of the method of Guarente *et al.* (11) to the F-IF gene. We describe the construction and identification of plasmids that direct the efficient synthesis of two proteins. The primary sequences apparently correspond to the sequence of F-IF in one case and to that of pre-F-IF in the other. F-IF produced in bacteria prevents viral growth as assayed *in vitro*.

MATERIALS AND METHODS

DNA Constructions. All techniques were as described by Guarente *et al.* (11). pTR56 (see Fig. 1) was constructed in two steps as follows. First, a plasmid (pLG111) was constructed that bears a *Hind*III synthetic linker three nucleotides before the ATG of pre-F-IF (6). This was accomplished by joining four DNA fragments: (i) a *Bam*HI-*Pst* I backbone fragment from pLG300 (11); (ii) a *Hind*III-*Bgl* II fragment containing the entire F-IF coding sequence from TplF319-13 (6); (iii) a *Pst* I-*Pvu* II fragment from pGL101 bearing the 5' end of the *amp* gene; and (iv) a *Hind*III linker. Ligation of these fragments fuses two complementary sticky ends (*Pst* I-*Pst* I and *Bam*HI-*Bgl* II) and two blunt ends (*Pvu* II-*Hind*III linker and *Hind*III linker-*Hind*III). The *Pst* I joining thus reconstitutes *amp*. Second, pTR56 was constructed by joining three fragments: (i) a *Pst* I-*Pst* I fragment from pLG111 bearing the 5' portion of *amp* and the 5' portion of the F-IF gene; (ii) an internal fragment of the F-IF gene extending from the *Pst* I site to the first downstream *Hinf* site (6), the latter having been rendered flush by DNA polymerase I (12); and (iii) a *Pst* I-*Bam*HI fragment from pLG300 (11) that bears the 3' end of *amp* and a 3' fragment of *lacZ*. The *Bam*HI end of this fragment had been rendered flush by DNA polymerase I (12). Ligation of these fragments generates two *Pst* I-*Pst* I fusions, one of which reconstitutes *amp*, as well as a *Bam*HI (filled in)-*Hinf* (filled in) fusion joining the 5' portion of the F-IF gene in phase with the 3' portion of *lacZ* (cf. refs. 6 and 11).

Radiolabeling of Proteins. The procedure has been described (11). Pulse labeling was with 300 μ Ci (1 Ci = 3.7×10^{10} becquerels) of [³⁵S]methionine and chasing was achieved by adding a 1000-fold excess of unlabeled methionine. This was done at 30°C. Labeled extracts were run on 15% acrylamide gels for analysis as described (13).

Preparation of Bacterial Extracts. Extracts were prepared essentially as described by Nagata *et al.* (14) except phenylmethylsulfonyl fluoride and EDTA were added in the lysis

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: F-IF, fibroblast interferon; SD, Shine-Dalgarno.

* The work described herein was carried out by the authors at The Biological Laboratories, Harvard University.

buffer at concentrations of 10 $\mu\text{g}/\text{ml}$ and 6 mM, respectively.

F-IF Assay. F-IF activity was determined by the cytopathic effect (CPE)-inhibition method essentially as described (2) with international standard F-IF (National Institute of Allergy and Infectious Diseases) as reference.

RESULTS

Method of Gene Expression. We desired to abut a DNA fragment bearing our portable promoter to the ATG encoding the amino-terminal methionine of pre-F-IF in one case and to the ATG encoding the amino-terminal methionine of mature F-IF in another case. This promoter fragment encodes the promoter and leader sequence of the *lac* operon but terminates two base pairs before the ATG that encodes the amino-terminal methionine of β -galactosidase (8–10). The mRNA encoded by the fragment does contain those bases called the Shine–Dalgarno (SD) sequence that, in conjunction with a properly positioned AUG, promote ribosome binding and initiation of protein synthesis (9–11). We have called the combination of the SD sequence of *lacZ* with the AUG of another gene a “hybrid ribosome-binding site,” and we have constructed various hybrid ribosome-binding sites to express a variety of prokaryotic and eukaryotic proteins (8–10). Although it is clear that the SD–AUG separation must be small (3–11 bases), we cannot yet predict the optimal placement nor do we know the effect of base sequence (15).

Our strategy, in outline, was as follows. Following the strategy of Guarente *et al.* (11), we fused a 5' portion of the F-IF gene with a large 3' fragment of the *lacZ* gene of *E. coli* (Fig. 2). This *lacZ* sequence encodes enzymatically active β -galactosidase, and as indicated in the figure, it is fused to a portion of *lacI*. This fused gene, carried on a plasmid, was not expressed, however, because it lacks transcription and translation initiation sequences. We then opened the plasmid at a unique restriction enzyme site 5' to the F-IF gene and resected DNA to various extents by using exonuclease (Fig. 2 A and B). This enabled us to position the portable promoter at many points near and in the 5' region of the F-IF gene (Fig. 2C). Lactose-negative

bacteria were transformed with a mixture of these plasmids, and those (rare) transformants expressing high levels of β -galactosidase were identified on the appropriate indicator plates. Sequence analysis revealed that two such isolates bore the portable promoter adjacent to the ATG of pre-F-IF in one case (pLG104) and the ATG of mature F-IF in the other (pLG117)

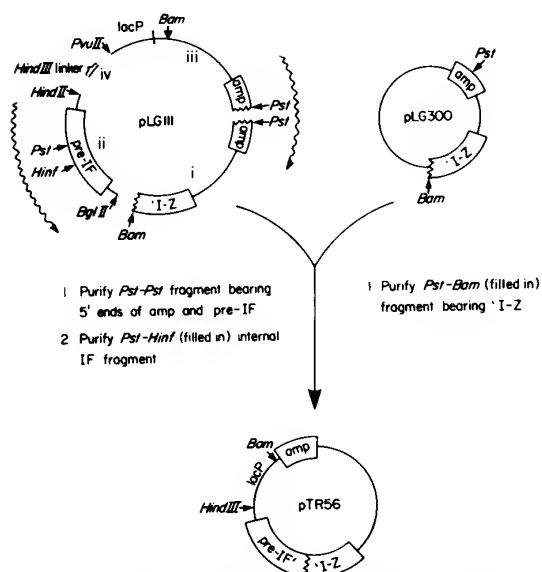


FIG. 1. Construction of pTR56. The arrows indicate the directions of transcription of the *amp* gene and the pre-F-IF gene.

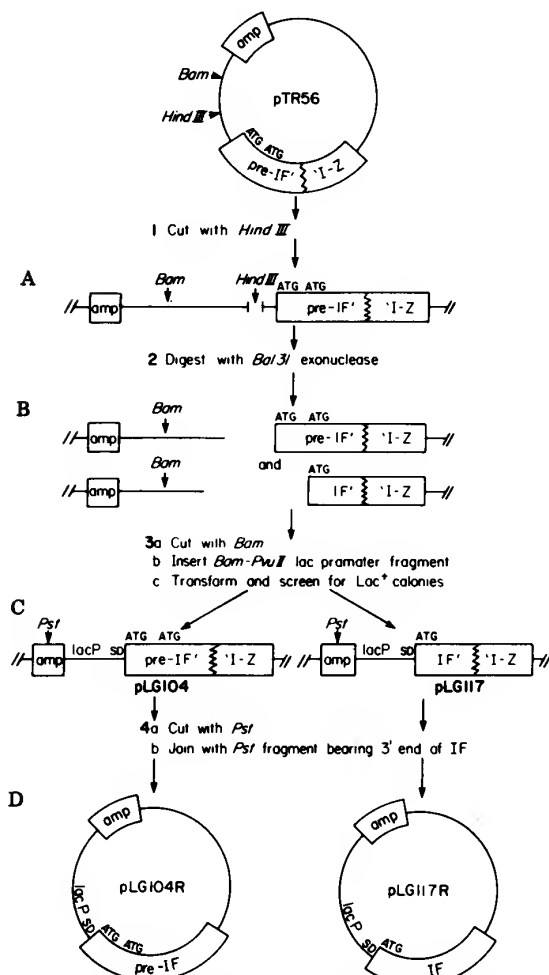


FIG. 2. Construction of plasmids that direct the synthesis of pre-F-IF (pLG104R) and mature F-IF (pLG117R). Plasmid pTR56 contains the 5' end of the F-IF gene fused to the 3' end of the *lacI*–Z gene and unique *Hind* III and *Bam* HI sites 6 and 116 bases upstream from the ATG of pre-F-IF. The ATGs encoding the amino terminus of pre- and mature F-IF are separated by 60 base pairs. Step 1, the plasmid was opened by cleavage with *Hind* III. Step 2, the linearized plasmid was digested for various times with the exonuclease *Bal* 31. Step 3, the resected plasmids were cut with *Bam* and religated with an excess of a fragment bearing the *lac* promoter that terminates 5 base pairs beyond the *lac* SD sequence. This 105-base-pair portable promoter is bounded by *Bam* and *Pvu* II ends. It is a derivative of the *lac* 95 portable promoter (see ref. 16). Strain LG90 [*F*[–], Δ (*lac*–*pro*)X111] was transformed with the ligation mix and was plated on lactose MacConkey indicator agar. Red (lactose-utilizing) colonies were picked, and the structures of two plasmids, pLG104 and pLG117, were determined. The relevant portions of these two DNA molecules are shown in the figure. Step 4, the genes for pre-F-IF and mature F-IF were reconstructed. Plasmids pLG104 and pLG117 were cut with *Pst* I, and the relevant portion was joined with a *Pst* I–*Pst* I fragment bearing the 3' end of the F-IF gene and the 3' end of *amp*. The designation R indicates the reconstructed F-IF-producing form of the plasmids.

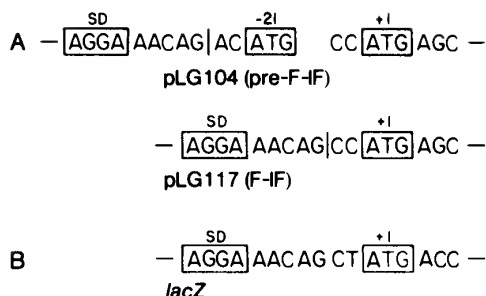


FIG. 3. (A) Nucleotide sequences of the DNAs around the regions encoding ribosome-binding sites of genes encoding pre-F-IF (in pLG104R) and mature F-IF (in pLG117R). The vertical lines separate sequences carried on the *lacZ* portable fragment from F-IF sequences. The boxes indicate the *lacZ* SD sequence and the ATGs encoding the amino-terminal methionines of pre-F-IF and of mature F-IF. (B) Corresponding region of wild-type *lacZ* (17).

(Fig. 2C; see also Fig. 3). The DNA containing the *lacZ* 3' gene fragment was removed and replaced with the 3' end of the F-IF gene, regenerating intact F-IF (pLG117R) and pre-F-IF (pLG104R), in which R indicates the reconstituted F-IF gene (Fig. 2D).

Plasmids that Direct the Synthesis of F-IF (pLG117R) and pre-F-IF (pLG104R). Fig. 3A shows the DNA sequence around the junctions of the portable promoter and the ATGs encoding the amino terminus of pre-F-IF (pLG104 and pLG104R) and that of mature F-IF (pLG117 and pLG117R). In each case, the SD sequence (AGGA) of *lacZ* carried on the portable promoter has been positioned seven base pairs from the ATG of F-IF. This is precisely the distance between the SD sequence and the ATG found in wild-type *lacZ* (Fig. 3B) (17). These particular placements were rare. In the screening that yielded pLG104, lactose-utilizing colonies appeared at a frequency of approximately 5%. Plasmid pLG117 was identified in a separate experiment involving more extensive exonucleolytic digestion. In this case, lactose-utilizing colonies were found at a frequency of only approximately 0.01%.[§]

Proteins Produced by pLG104R and pLG117R. The experiment of Fig. 4 uses the "maxicell" technique to display those proteins encoded by pLG104R and pLG117R. Suitably treated maxicells differentially incorporate radioactive amino acids into plasmid-encoded proteins that are easily visualized by autoradiography after polyacrylamide gel electrophoresis (11, 18). pLG104R and pLG117R each direct the synthesis of one protein in addition to β -lactamase. In the case of pLG117R, a protein was produced with a molecular weight of approximately 20,000, consistent with the predicted (from the DNA sequence) size of unglycosylated mature F-IF (6). In the case of pLG104R, the protein produced had a molecular weight of about 23,000, which corresponds to the predicted molecular weight of unglycosylated pre-F-IF.

Fig. 4 also shows the fate of pulse-labeled pLG104R and pLG117R proteins in a maxicell experiment. Densitometry tracing of the gel (not shown) reveals that pre-F-IF was com-

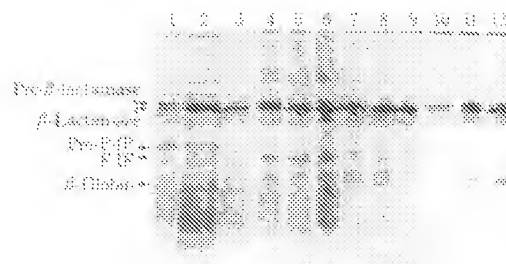


FIG. 4. Proteins encoded by plasmids pLG104R (pre-F-IF) and pLG117R (F-IF). Plasmid proteins were labeled by the maxicell technique (18). Proteins encoded by pLG104R were labeled for 5 min with [³⁵S]methionine (lane 1) and then chased with nonradioactive methionine for 12 min (lane 2) or 50 min (lane 3). Labeling and chasing were terminated by freezing the cells. After spinning for 2 min in an Eppendorf centrifuge, the cell pellets were suspended in Laemmli's sample buffer (13), incubated for 3 min at 100°C, then subjected to analysis by polyacrylamide gel electrophoresis as described (13). Similarly, proteins encoded by pLG117R were labeled for 5 min (lane 4) and chased for 12 min (lane 5) or 50 min (lane 6). Lanes 7, 8, and 9 display the same labeling protocol performed on plasmid pGL101, which encodes only β -lactamase. Plasmid pLG302-2R, which directs synthesis of rabbit β -globin (5000–7500 molecules per cell) (11) was likewise analyzed as shown in lanes 10, 11, and 12.

pletely degraded in a 50-min chase (lanes 1 and 3), and F-IF was about 50% degraded in that time (lanes 4 and 6). Although processing of β -lactamase from pre- β -lactamase is evident, pre-F-IF is apparently not processed.

We estimated the level of our F-IF and pre-F-IF synthesis in two ways. First, we measured the amount of radioactivity incorporated during a 5-min pulse into F-IF and pre-F-IF in a maxicell experiment. We compared these values with a known standard, namely, rabbit β -globin synthesized by the plasmid pLG302-2R (11) (see Fig. 4). This comparison suggests that, were the F-IF molecules stable, the steady-state levels would be 5000–10,000 molecules per cell. Second, we found that growing cells bearing plasmids pLG117 and pLG104 synthesize about 1200–1400 units of β -galactosidase (19). Assuming that the hybrid F-IF (or pre-F-IF)- β -galactosidase molecules have the same specific activity as β -galactosidase, this value represents 5000–10,000 molecules per cell (11). Previous experience with β -galactosidase hybrid proteins modified at their amino termini suggests that they are stable during growth (20).

Antiviral Activity of the Human Fibroblast Interferon Polypeptide Synthesized in *E. coli*. Extracts of bacteria bearing plasmid pLG117R inhibited growth of vesicular stomatitis virus on human fibroblast cells in a typical interferon assay (inhibition of cytopathic effect) (2, 21). This activity was abolished by antibody to F-IF but not by antibody to leukocyte interferon (Fig. 5). Moreover, extracts of bacteria carrying pLG104R, pLG115R, and pBR322 failed to manifest antiviral activity. The antiviral activity directed by pLG117R survives a sojourn at low pH (pH 2.0) or treatment with DNase and RNase, but it is abolished by trypsin treatment (data not shown). Assuming that unglycosylated F-IF is as active as glycosylated F-IF [2×10^8 units/mg (22)], the activity we typically recovered would correspond to approximately 50 interferon molecules per bacterial cell.

DISCUSSION

Our results strongly suggest that our application of the method of Guarente *et al.* (11) to the F-IF gene isolated by Taniguchi *et al.* (6, 7) has produced two plasmids (pLG104R and pLG117R) that direct the synthesis of proteins whose primary sequences correspond, respectively, to those of pre-F-IF and F-IF.

[§] This screening also yielded a third fusion, pLG115. In this case, the portable promoter is abutted to nucleotide 18 of the sequence encoding the F-IF leader some 20 bases from the nearest possible initiator triplet (6). Experiments similar to the experiment of Fig. 4 suggest that, in this case, protein synthesis initiates at the internal ATG located at position 175 in the DNA sequence published by Taniguchi *et al.* (6, 7). This ATG is fortuitously preceded by an SD-like sequence in the F-IF gene. We do not understand how the portable promoter placement in pLG115 enhances the efficiency of utilization of this ATG.

